THEORY OF EXPERIMENTAL INFERENCE

THE MACMILLAN COMPANY
NEW YORK · BOSTON · CHICAGO · DALLAS
ATLANTA · SAN FRANCISCO

MACMILLAN AND CO., LIMITED
LONDON · BOMBAY · CALCUTTA MADRAS
MELBOURNE

THE MACMILLAN COMPANY
OF CANADA, LIMITED
TORONTO

THEORY OF EXPERIMENTAL INFERENCE

C. West Churchman
WAYNE UNIVERSITY

New York
THE MACMILLAN COMPANY
1948

PRINTED IN THE UNITED STATES OF AMERICA

To
Ernest R. Rechel

Preface

This essay is really an attempt to preach the gospel of modernism to two groups; in this dual purpose it may well fail because of the duality itself. But the attempt had to be made in this way, for the general moral is that these two groups, the philosophers and the scientists, can no longer live apart. The essay has attempted to preach to the modern philosopher of science that we can no longer conceive of scientific method in the older empirical fashion of "discrete" observational sentences. Modern science, and by this I mean the science we have inherited from the nineteenth century, has forced us to recognize that no question of science can have meaning unless it involves a problem of measurement along a scale whose order is that of the continuum. The consequence is that the "verification" of answers to questions must become an "ideal" in the sense to be discussed herein. The approach to this ideal requires that we put far more into the process of question-answering than is contained in so much of present-day philosophical writing. The methodolgy of science can no longer be conceived as "empirical" in the traditional sense of the term.

As the essay will try to show, the presuppositions of inquiry become far more complicated than early experimental science dreamed. This is emphasized by the persistent claim of the essay that the simplest question of fact in science requires for even an approximation, a judgment of value. This is a far stronger claim that even the contemporary pragmatic writers are willing to make for ethical theory. We are *not* merely claiming that ethical judgments can be included within the scope of science; this claim is by now well recognized by serious students of method. We are rather making the

much stronger claim that the science of ethics (like all the principal branches of science) is *basic* to the meaning of any question the experimental scientist raises. All the so-called "facts" of science imply for their meaning a judgment of value.

To the scientist, the essay has attempted to preach the moral that the process of raising a question and proceeding to reply to it, is a process which eventually demands the active cooperation of all fields of research. We can no longer visualize the experimental scientist as the lonely worker who designs and completes his own researches. This lesson is all the easier to make in view of the extensive work done by the mathematical statisticians during the past half century. The statisticians have shown very clearly that even to *frame* a sensible question within science requires considerable research; they show that the problem of the most efficient use of the observations requires an extensive cooperation between methodologist and specialist. But the lesson preached herein goes beyond that offered by the statisticians. We have attempted to show that besides certain statistical considerations, the self-conscious experimenter must take into account very general problems concerning the natural universe within which he is solving his special problems.

As might be expected, some of the ideas developed in this essay suggested others along the same lines, and have consequently led to more extended treatments. This is especially true of the psychological concepts discussed in Chapter XIV, and the social and ethical concepts discussed in Chapters XV and XVI. These more extended treatments will appear in a work prepared with R. L. Ackoff entitled "Psychologistics" which is now in preparation. The present essay is devoted to the problem of methodology in general, while the other work considers the problem with respect to its application to the social and ethical sciences.

Anyone familiar with the works of E. A. Singer, Jr., cannot fail to find his influence throughout. It would be more economical to enumerate the points of view that are not directly his, than to enumerate those that are: e.g., the classification of the schools (though modeled after Singer's *Dialectic of the Schools*, the present treatment introduces the relativist school and emphasizes the con-

flict between method and certainty in science); the treatment of the approach to answers of questions of fact (phrased differently, though with the same intention as Singer's); the treatment of the modern theory of statistical tests throughout; the chapters on personality, the social group, and the role of the science of ethics. T. A. Cowan has played a very important part in the writing of this work, since much of it has come out of conversations with him. R. L. Ackoff has helped edit the entire work, and the material in the chapters on personality, the social group, and the science of efficiency is a result of cooperative work with him.

To my friends at this University, and at Frankford Arsenal, there is owed a large debt. The latter have shown an astonishing interest in these fundamental matters, and their cooperation is really the stimulus of this essay.

The material in Chapter IV, and in Chapters X–XII originally appeared in a somewhat different form in the *Journal of Philosophy* and the *Philosophy of Science*, respectively; permission to re-use this material has been kindly granted by both journals.

I am endebted to the Faculty Research Committee of the University of Pennsylvania, for a grant for the mimeographed publication of the first version of this essay. Many helpful comments were received on this earlier version from scientists and philosophers.

C. W. C.

Contents

Chapter I On the Nature of Statistical Tests

We can best characterize the sort of problem with which we are here concerned by means of an illustration. Suppose, as a laboratory technician, you are faced with the problem of deciding whether a new product you have developed is any better than the one now being made. We will suppose that your decision is to be based upon two comparable sets of data, the first representing the amount of force just necessary to break five samples of the new product, the second the amount of force just necessary to break five samples of the regular. Your results run as follows:

Regular	*Special*
5.01	5.62
4.76	5.25
4.98	6.31
5.56	5.07
5.05	5.42

Has there been any significant improvement in quality; i.e., should the laboratory offer the new product as a definite improvement, worthy of such expense as might be involved in the changeover?

The intuitive enthusiast might answer the question by pointing out that a "glance at the data" makes the whole matter "obvious"; all but one sample of the regular was poorer than the worst of the special, and one can see without studying the matter that the new development is a significant improvement over the old. The antithetical reply of the cautious sceptic would consist in pointing out the wide dispersion in the results, and the relatively high value of one of the regular samples; as a consequence he would refuse to make any inference from the data.

It is the avowed purpose of modern statistical tests to avoid the confusion to which both these intuitive replies lead. Setting aside the enthusiast's appeal to the obvious, is his conclusion really sound? He asserts that the special product must be better because in all the samples tested it was better than any of the regular, except one. Is this cogent reasoning? Could not such a result occur by chance even when no real difference exists? On the other hand, is the sceptical answer sound in asserting that the readings are too widely dispersed to make any sensible inference?

In general, what does constitute sound reasoning from data of this sort? How does one know whether the inference one makes is correct?

The modern statistician recommends an "exact" procedure to be followed in these matters, which he claims constitutes the sound reasoning that is required, provided certain general assumptions hold. We shall want to see first just what this procedure is, and then to examine the statistician's grounds for asserting that his procedure is in some sense the "best." In fact, we may make such a study the basis of this work: to examine critically the statistical claim to exactness and validity in matters of inference.

In the case of the illustration above, the statistician offers the following as a possible procedure:

1. Calculate the average, $\bar{x}$, of each set of readings. In mathematical notation

$$\bar{x} = \frac{\sum^n x_i}{n}$$

where $\sum^n x_i = x_1 + x_2 + x_3 + \cdots + x_n$, and x_1, x_2, x_3, $\cdots$, etc., are the individual readings. n is, of course, the sample size. Let us call $\bar{x}_1$ the average of the regular samples, and $\bar{x}_2$ the average of the special. We have

$$\bar{x}_1 = 5.072$$
$$\bar{x}_2 = 5.534.$$

2. Now calculate the following quantity for each set of data:

$$s^2 = \frac{\sum^n (x - \bar{x})^2}{n - 1}.$$

This value is extremely useful in statistical matters, and is called an unbiased estimate of the "variance" of the readings, the variance being a measure of the dispersion of the results; it is designed to replace vague and intuitive guesses as to the scatter of a set of data. Just why we use this measure of the dispersion must await our examination of the validity of the entire method.

In this case, we have

$$s_1^2 = .0871$$
$$s_2^2 = .2296.$$

3. Now calculate the value t, where

$$t = \frac{(\bar{x}_2 - \bar{x}_1)(n)^{1/2}}{(s_2^2 + s_1^2)^{1/2}} = \frac{(.462)(5)^{1/2}}{(.2296 + .0871)^{1/2}} = 1.83.$$

4. Using this value of t, we are required to consult tables of "Student's t-distribution," to be found in any standard statistical textbook. In order to use these tables one has first to calculate a quantity known as the "Degrees of Freedom" (D.F.), which is given by the simple equation:

$$\text{D.F.} = 2n - 2 = 8.$$

The tables are constructed to give us a certain probability; in order to understand the meaning of this probability, imagine yourself to have divine insight into the truth of these matters, and to be picturing the laboratory technician repeating the above test on five samples from each product, over and over. Suppose also that the true long-run quality of each product is exactly the same. Then, we now ask ourselves, in this repetition of the experiment, how often will the technician obtain a value of t as large as or larger than 1.83? Put in more prosaic language, just how improbable is a value of t of 1.83 or larger for samples of five, when no real difference exists?

The tables are designed to give the probability of obtaining such a difference when no real difference exists. Statisticians view the matter from the logical side somewhat as follows. Let us call the proposition that in the long run no real difference exists the "null hypothesis"; following the modern inclination to symbolize, let us use the symbol "H_0" for the null hypothesis. Then Student's tables

are designed to give us the probability that $t \geqq 1.83$, *if* H_0 is true. In this case, the probability is about .1; that is, the observed difference could occur about one time in ten under the assumption that no real difference exists. This is not very improbable, and hence in general one is not inclined to reject H_0.

Please note, the statistician hastens to add, that one is not forced to accept H_0 either. Perhaps the result given by the statistical analysis is that not enough observations have been taken. To say that, because the data do not reject H_0, they must lead to its acceptance is as bad as saying that since our present knowledge does not lead us to reject the statement that there is life on Mars, we must accept the statement.

The curious mind is naturally led to inquire into the reasoning behind the statistician's method, and indeed such inquiry is essential if we are to lay any claim to the preference of his method over the purely intuitive one.

In the first place, it is to be noted that the above procedure is not recommended in general, and without some prior knowledge. Specifically, the procedure is to be regarded as "sound" if the following conditions hold:

1. The samples are each drawn from a "normal" universe.[1]
2. The samples are drawn at random.[1]
3. The true variances of the universes are the same, i.e., the data from each sample are subject to the same degree of dispersion (though the exact magnitude of this degree is unknown).

It begins to look now as though the statistician's procedure had a long story behind it; the four steps outlined above were really the final steps in a much longer study designed to confirm these three statements. In order to understand this background, we need to gain some acquaintance with the new statistical terms it introduces.

1. The meaning of a normally distributed universe may be derived intuitively by picturing a gigantic bowl which holds an indefinitely large number of chips. Each chip is numbered, with either a positive, negative, or zero quantity. If the potential draws from the

[1] For the mathematical definitions of normality and random, see next page.

bowl are to represent a normal universe, then the following conditions must hold (though these are not sufficient): a) the chance of drawing a negative number of a certain magnitude is equal to the chance of drawing a positive number of the same magnitude (the distribution is *symmetrical*), b) the chance of drawing a given number becomes less and less the farther such a number deviates from the value zero, and c) the chance approaches zero of drawing very large deviations. The zero chips may be taken as the long-run mean of a series of observations, and the properties of a normal distribution of observations are that plus deviations from the mean are in the long run as common as negative, and that very large deviations are very rare.

The statistician is naturally forced to use a mathematical expression to characterize the normal universe. He does this by giving a formula for the probability P of drawing observations that deviate from the mean by no more than the quantity z:

$$P = \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-z}^{z} e^{-t^2/2\sigma^2}\,dt$$

where σ^2 is a measure of the dispersion of the observations (called the "variance"). The value σ^2 is closely related to the value s^2 we computed above, and as a matter of fact, is the basis of our computing s^2 in the first place. The value of σ^2, like the long-run value of $\bar{x}$, is not known as a result of only a few samples, but it may be estimated; in fact, the quantity s^2 represents a "good" estimate of σ^2. The equation for P given above defines the universe, and is known as the elementary probability law of the universe.

2. The concept of "randomness" is much more difficult to characterize. Intuitively, a sample is said to be drawn "at random" if one can find no regular way of ordering the observations. For example, if each observation is less than its predecessor, or if every odd observation is higher than its succeeding even observation, the set is not taken to be random. A formal definition of "random sample" consists of the following necessary and sufficient conditions: (1) the elements of the sample are drawn from the same universe, in the sense that they obey the same elementary probability law and (2) the elements are *independent* draws, in the sense that if P_1

is the probability that a certain "draw" or "observation" x_1 lies in an interval, I_1, and P_2 is the probability that x_2 lies within an interval I_2, then the probability of *both* occurrences (x_1 in I_1, x_2 in I_2) is exactly P_1 times P_2. This is a *formal* definition of randomness, we say. But the pertinent question as far as the experimenter is concerned is *nonformal:* how in practice can one verify these conditions of randomness?

It is important at present, however, that we have some idea of the statistician's motive for insisting on randomness, for this concept has always been the most difficult to define and characterize in practice. The reasoning behind his insistence is actually the whole reasoning behind scientific methodology, and we can only make a beginning at this point in our understanding of it. In a sense, randomness represents the "freedom" of observations; if the occurrence of any observation followed a definite and prescribed pattern, then by discovering this pattern we could discover all we need to know, and modern statistical techniques would be outmoded. Imagine, for example, a man spinning a coin; if he regards the set of spins as random, then his method of predicting a head or a tail will depend on his observations, but if he regards the set of spins as determined a priori by some known fixed set of laws, then no spin is *random*, and the prediction of a head or a tail does not depend on his past observations. Hence randomness really dictates the method we are to use in making predictions, and if randomness does not exist, then statistical techniques of handling data are not in general applicable. The statistician's claim to usefulness lies in the fact that when the experimenter faces nature, he rarely if ever faces it with a complete a priori knowledge of what is going to happen, and in most cases, if not all, his observations have a random character.

Thus randomness forms the background of statistical tests, and these tests are all of them based on the hypothesis that certain of the observations are not a priori determined for the observer. This, we have said, forms the "intuitive" justification of randomness, but we shall see that it raises some rather awkward questions in connection with scientific methodology in general. For one thing, such a description seems to imply an indeterminism in nature of some sort, an indeterminism that is forever unconquerable. These

matters and others connected with the concept of randomness we shall reserve for later chapters.

The questions that we wish to raise in connection with this explanation of the statistician's terms are two: (1) why do the observations have to satisfy these rather complicated conditions, and (2), if they do have to, how do we know they do?

The answer to the first of these is entirely formal, and depends upon mathematical considerations. The argument runs somewhat as follows: suppose observations to be made at random from some normal universe. Then we can derive mathematically the "distribution function" of a certain statistic "t"; this statistic is calculable entirely in terms of the observations. If the observations are divided into two sets of n each, as they are in the example above (regular and special), and if we calculate $\bar{x}$ and s^2 for each set, then t is defined, as above:

$$t = \frac{(\bar{x}_2 - \bar{x}_1)(n)^{1/2}}{(s_2^2 + s_1^2)^{1/2}}.$$

The "distribution function" of t is analogous to the distribution function for a normal universe, and from it we can derive the probability of finding a value of t in any interval, and in particular the probability that a value of t will be as large as or larger than the one observed. The distribution function of t, under the hypothesis that the two samples are drawn at random from the same universe, was discovered first in one form by Helmert, in 1876, and later by "Student" in a classical paper appearing in 1908 (3).[1] If H_0 holds, then in the long run (i.e., for infinite sample sizes), $\bar{x}_1$ will equal $\bar{x}_2$ and t will have the limiting value of zero; the distribution function of t gives us the probability of observing as large a deviation from zero as the one given, provided H_0 is true. As n, the sample size, increases, the chance of observing a large deviation decreases towards zero; or, if the observations are very accurate, and s_2^2 and s_1^2 are very small, then the chance that $\bar{x}_2$ deviates much from $\bar{x}_1$ (under the supposition that H_0 is true) is very small. These intuitively sound properties of t can be derived from the conditions we

[1] Numbers appearing in parentheses refer to the numbered references at the end of each chapter.

have discussed previously, and this is actually our basis for wanting the observations to satisfy normality and randomness. If they do satisfy these conditions, then the rest of the matter becomes quite simple. We can derive rather exact conclusions that are of extreme importance in making decisions.

As a matter of fact, the example above is an illustration of what the statistician calls the "uniformly most powerful test" of an hypothesis. He views the matter as follows: there are, after all, an infinite number of methods the laboratory technician could have used to decide one way or the other. He could have ignored all but the largest observation in each sample, and based his decision on these two values; or he could have compared the averages $\bar{x}_1$ and $\bar{x}_2$ with respect to the "mean deviation." If he used the mean deviation, he would add the deviations of each reading from the mean, irrespective of sign, and divide the total by n. Thus, he would obtain

$$\text{mean deviation for regular} = .195$$
$$\text{mean deviation for special} = .325.$$

He might then argue that a real difference must exist, because the difference in means, .462, exceeds the larger of the mean deviations. He would certainly be proposing a method which one could follow in every case, and in so doing he raises the question whether his method is not really sounder than the one proposed above.

To show that this alternative method is not as "good" as the other, statisticians frame the following argument. In view of the uncertainties associated with the taking of observations, it must be admitted that however we decide an issue on the basis of the evidence at hand, we always run two risks, provided we make any kind of positive assertion: (1) we may reject H_0 when it is really true (this is called a type I error, or an error of the first kind), or (2) we may accept H_0 when it is really false, i.e., when some other possible alternative hypothesis is true (type II error, or error of the second kind). In a simple case of two alternative hypotheses, this situation reduces to that of accepting H_1 when H_0 is true (type I), or accepting H_0 when H_1 is true (type II). Now it seems evident enough that the "best procedure" will be one that minimizes

both types of error; that is, one procedure is better than another if both types of error are smaller in one case than they are in the other. We can feel perfectly safe in recommending that procedure for which the risks of committing an error are as small as possible, and this will be taken as the grounds for preferring one method over another.

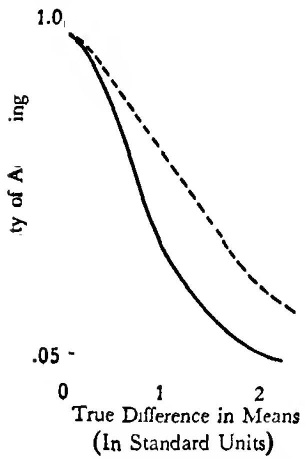
This simple view of the matter, like most simple views, is not exact enough to provide a good basis for evaluating one method over another. For one thing, the term "reduce both types of error to a minimum" is ambiguous. Suppose one method involves a larger type I error but a smaller type II error than another method; which method is preferable? As an extreme case, suppose someone simply asserts that he is always going to accept H_0 no matter what the observations may be; thus a research worker might become so confident that a new product was not an improvement that he refuses to pay any attention to comparative tests. In this case, his chance of rejecting H_0 when it is true, and hence his type I error, is zero. With respect to this type of error, therefore, his method must be at least as good as any other, and is certainly better than the one we have proposed above. His type II error, on the other hand, will be larger.

In view of this situation, a more precise formulation of the statistician's criterion of a "best" method is needed. This has been suggested in a series of important papers by Neyman and Pearson (1, 2).

Suppose that α is any given type I error, or probability of rejecting H_0 when it is true; then, say Pearson and Neyman, the best method will be that method which minimizes the type II error, regardless of what the alternatives to H_0 may be. That is, the best method reduces type II errors, independent of what the alternatives may be. This phrase, "independent of the alternatives," is important, for it means that the best method does not depend upon which alternative hypothesis to the null hypothesis actually holds.

This manner of stating the criterion of "best" statistical test avoids the paradoxical aspects of the illustration we have cited above; the method of always accepting H_0 when it is true is not even comparable to other methods unless α happens to be zero.

But in this case, the method does not minimize the risk of committing a type II error. If α is zero, the "Student t-distribution" would never reach a decision with only a finite set of observations, and hence would never commit either type of error. On common sense grounds, in this limiting case the Student-method seems preferable, for no one could ever assert absolute confidence ($\alpha = 0$) on the basis of a finite set of observations.



This matter may be clarified graphically as follows. The accompanying figure represents an "operating characteristic curve." It is designed to show the probability of making an error of either type; along the abscissa there is recorded the difference between two means (in so-called "standard units," i.e., in units of the standard deviation). The ordinate shows the probability of accepting the null hypothesis (no real difference) under a given statistical test, when the *true* state of affairs is that indicated on the abscissa. Thus, for the solid line shown, the probability is .05 (one chance in twenty) of accepting H_0 when the true difference in means is as large as 2. The *ideal* statistical test would be zero along every point of the abscissa except the origin, where it would be 1. This ideal is never attained for any finite set of data; instead, we must compare the existing methods to determine the best available. Thus, the statistical test designated by the solid line is better than that designated by the dotted line, because, for a given probability of rejecting H_0 when it is true, the first always gives a lower chance of accepting H_0 when it is false. There are tests (such as the Student t-test) which are better than any other possible test, no matter what value we read along the abscissa.

Pearson and Neyman have found it impossible to apply the criterion of a uniformly most powerful test in all cases. It may be applied to the example given at the beginning of this chapter, provided the observations within a sample are random draws from

the same normal universe. In this case, as we have indicated, the Student t-test is the best, or most powerful, test that can be found; any other method, for a fixed type I error, will yield at least as large a type II error (actually larger in all but the trivial case of one observation). Thus the method of using the mean deviation, suggested above, is not as "good" as the Student test, on the basis of the Pearson-Neyman criterion.

But there are examples of a more complicated sort in which no "best" test in the sense described above can be found; in fact, it may be shown mathematically that in these instances no "best" test actually exists. In these cases, the criterion of best has to be modified; for example, a method may be the best if it satisfies the conditions given above "in the limit," i.e., as the sample size becomes larger and larger. Such tests are called "asymtotically most powerful tests of statistical hypotheses," and are often taken to be preferable to all others.

The difficulties involved in modifying the criteria of "best" test when no best test exists need not detain us here. The basic principle behind all such modifications is that for a given type I error, the type II error should in some general sense be minimized. The chief result of these modern studies into statistical method, from our point of view, is that they attempt to provide an exact answer to the question we posed above: what forces the statistician to insist on rather complicated conditions that the observations must satisfy? The answer can now be made quite simply: unless we make certain universal presuppositions about the observations, we cannot determine which is the best or most powerful method of reaching a decision. And, it might be added, without our having criteria for the best method, then the whole of science becomes a haphazard process, without a precise methodology.

The situation is analogous to that which arises in formal science; in order to deduce theorems in geometry, we require certain principles as criteria of correct or "best" deduction. If such criteria do not exist, then any result can be obtained, and the science becomes meaningless. In just the same manner, unless we have exact criteria of correct tests in experimental work, then our whole method has no foundation, and any result is as good as any other.

It should be noted in passing that we do not have to make exactly those presuppositions we did above, with respect to the observations. For example, instead of a normal universe, we might have used the simpler presupposition that the observations came from a universe with a continuous distribution function; the statistical test would not have been the same, but we could still find a "best" test. Indeed, an important branch of modern mathematical statistics is devoted to an attempt to find the criteria of best tests when we make simplified presuppositions about the universe of the observations (the so-called "nonparametric" problem).

But however simplified the presuppositions may become, *some* presuppositions must be made, and the question we asked earlier is still pertinent: how are we to know whether the observations really satisfy the presuppositions?

Has the statistician actually made any advance over the intuitive method? Are we any better off by simply pushing our "guesswork" back a few steps? We may now be provided with an exact method for determining whether or not the special is better than the regular product, but is this really any help, when it turns out that this method rests on certain presuppositions which we cannot confirm? What good does it do to define so carefully the criteria of a "best" method, when such criteria are only valid provided doubtful conditions hold?

To the difficulties involved in confirming these presuppositions is added another, perhaps more subtle one. Certainly it is frequently overlooked by the modern statistician. We say that a type I error is committed by rejecting H_0 when it is "true," and we base our criterion of best on this concept. But what does "true" mean in this context? Is it fair to leave the matter up to the "common sense" of the reader, and hope that everyone will have some sort of "mutual agreement" on this matter? The history of philosophy provides a cumulation of evidence against any view that all men mean the same thing by "truth." To some, truth is a "correspondence with reality," to others it is the intuitively self-evident universal proposition, to others it is what is to be found in or implied by the Koran, to others it is what the senses tell them.

The matter is not trivial. Quite the opposite. If modern statis-

tics is to claim any soundness in its reasonings concerning most powerful tests, it must eventually come to explain what the concepts "true" and "false" mean, for these are basic to an understanding of the type I and type II errors, and these in turn are basic to the entire theory.

Of course, "true" and "false" may receive a purely mathematical definition, as they do at the hands of Pearson and Neyman. The authors use a geometrical analogy, and the "true" and the "false" define regions in an n-dimensional space. But it is not this sort of formal defining that interests our laboratory technician; he wants to know about truth in a much more "practical" sense. In the example above, he would say he had made an error if his prediction "runs him into trouble" later on. The task that faces us is to characterize as exactly as possible the meaning of "running into trouble," i.e., to define truth and falsity on nonformal grounds.

Before proceeding to formulate answers to these questions, it will be advisable to generalize upon the example given above, and to present a formal picture of the method of making a response to a question by the techniques of mathematical statistics. Since we are later to assert that all questions receive responses in this manner, the technique outlined in the next chapter is to be regarded as a general account of the method of making experimental inferences. The fundamental questions this account raises will be the subject matter of the subsequent chapters.

REFERENCES

1. Neyman, J., and Pearson, E., "On the Problem of the Most Efficient Tests of Statistical Hypotheses," *Philosophical Transactions, Royal Society*, Series A (1933).
2. Neyman, J., and Pearson, E., "Contributions to the Theory of Testing Statistical Hypotheses," *Statistical Research Memoirs*, Vol. 1 (1936).
3. Student, "On the Probable Error of the Mean," *Biometrika* (1908).

Chapter II General Methodology of Inference

In this chapter there is presented an outline of the general method of responding to questions in accordance with the procedures discussed in Chapter I. The purpose of such an outline is to enable us to ask specific questions of a basic sort about the methodology of science. It will turn out that a response to even the simplest question involves a rather complicated method, and this consequence will no doubt raise in the reader's mind a question concerning the generality of the method: whether all possible meaningful questions must be treated in this manner. The discussion of this problem, or rather the discussion of a generalized version of the problem, we leave for later chapters.

We now suppose that a question has been asked, and that a response is to be found by taking a series of observations. We wish to examine the logic of making an inference from such observations, this logic being based on the developments that have been made in mathematical statistics during the past half-century.[1]

In effect, we assume that the most precise available methods of

[1] A list of references is appended to this chapter; as before, citation of the literature throughout this chapter has reference to the numbers of this list. Some general remarks about these references may be useful to the reader. General texts which explain certain statistical techniques and provide a general insight into the logic of the methods are (8, 9, 12, 16, 18). The theory of "best" methods and the like is discussed in (1, 2, 7, 13, 14, 19, 20). Problems of general interest connected with special methods are discussed in (15, 17, 21). An elementary discussion of the presuppositions involved in a very widely used method, the "analysis of variance," are discussed in (5). Elementary mathematical theory is given in (10); advanced mathematical theory in (2, 4, 11, 22). Some risk curves associated with widely used methods are given in (6). A more formal treatment of the material in this chapter is given in (3). Needless to say, the references are by no means complete on the various subjects cited.

drawing inferences from data are those developed by the statisticians, and that the best course to be followed by the philosopher interested in experimental inference is to examine these methods. This is an assumption that no doubt will eventually need some defense, but for the present we take for granted its validity as a basis for a beginning.

The following analysis, it should be understood, differs from the usual formal presentation of statistical tests of hypotheses (19) in that its aim is *not* to define the most general problem of statistics, but rather to define the general problem of experimental methodology. In this sense, the usual presentation is formal and mathematical; the following is formal and philosophical.

We suppose our task to be one of formulating responses to such questions as: "How hard is this metal?" "What is the sensitivity of this organism?" "What is the law of stability of a certain compound?" The following are taken to be necessary conditions for formulating responses:

1. *A formal theory of probability must be given.* The term "formal" is meant to imply that the theory must comply with the demands of an axiomatic system: there must be a set of undefined concepts, a set of formal axioms concerning the undefined concepts, a set of rules for constructing theorems, and a method for deciding whether or not the set of axioms is "consistent" (with the understanding that the rules and the consistency-proof are not necessarily expressed in the same language as the axioms).

The exact nature of the probability theory cannot be stated in rigid terms, but modern statistical theory has demanded that at least the following concepts be defined or be taken as undefined within the theory:[1]

a) "Event," or "observation." The events (or observations) may be defined as singular propositions of the form "The specified coin fell heads on the *n*th toss," or "The *n*th reading was 3.65 cm." Or, as is more frequently done in statistical theory, the events may be taken as numbers, i.e., quantitative elements. It is to be noted that

[1] The reader unfamiliar with certain logical and mathematical aspects of the following may wish to refer to the summary at the end of the discussion of probability theory, p. 19.

in the latter case the events may be ordered *pairs*, or *n*-tuples, of numbers; for example, we may be observing the way in which a body moves along a straight line in time, so that an "event" is a number pair, position versus time.

b) "Sample," which may be regarded in several ways. If the events (or observations) are defined as singular propositions, then the sample is regarded as a *logical conjunction* of such propositions. The number of *logically* [1] independent elements of such a conjunction is called the "sample size," and corresponds to the "number" of observations or events. On the other hand, if the events are taken as numbers, then the "sample" is commonly regarded as a point in an n-dimensional space, the location of the point along the ith coordinate axis being determined by the magnitude of ith observation. Either definition of sample is adequate in general, though the latter does permit certain simplifications of the subsequent procedure. If the "events" are given as ordered n-tuples, then the meaning of a "sample-point" must be generalized accordingly.

Note that the definition of sample is *formal* in the sense that we have not given a sufficient basis for an experimenter to decide whether or not a certain type of behavior on his part necessarily yields a sample. In other words, so far a sample is only an aspect of a formal probability theory. This remark, of course, applies to the other terms we are about to define as well.

c) "Elementary probability law," which has been defined in the previous chapter as the law which states the probability of an observation's falling within any specified interval. This probability may be regarded as the limit of the long-run frequency with which observations would fall within an interval, if the number of observations were increased without limit. It should be noted that the law has quite general application. There are cases, for example, where the events or observations can take on only discrete values (e.g., the integers, or a finite set of the integers). This case corre-

[1] Throughout this chapter, we must make use of the term "independence" in two quite distinct senses, the *logical* and the *statistical*. The former refers to propositions in general, a set of propositions being logically independent if one cannot be derived as a consequence of any combination of the others. The latter is discussed under paragraph e) below.

sponds to the "statistical" case in classical error-theory.[1] However, in a very wide number of cases, the elementary probability law can be expressed by an integral, of the form

$$= \int_a^b p(x)\,dx$$

where P is the probability that an observation occurs between a and b. $p(x)$ is then called the elementary density function or "distribution function" of the observations. In the case of "normal distribution functions," which are most frequently presupposed, $p(x)$ has the form

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-(x-\mu)^2/2\sigma^2}$$

and, when plotted, appears as a bell-shaped curve, which approaches the axis asymtotically in either direction. The values μ and σ^2 completely define $p(x)$ and are called the "parameters." Now, neither μ nor σ^2 is known in practice, and an important problem is often to find a "best estimate" of these unknown quantities. μ and σ^2 are commonly called the (true) "mean" and "variance" respectively, because, in general, we have as definitions

$$\text{(true) mean of } x = \int_{-\infty}^{\infty} x p(x)\,dx$$

$$\text{(true) variance of } x = \int_{-\infty}^{\infty} x^2 p(x)\,dx - \mu^2.$$

when the integrals exist, and it may be shown that for the normal distribution, μ and σ^2 satisfy these properties respectively. The square root of σ^2, σ, is usually called the (true) "standard deviation."

When the observations are ordered n-tuples, we obtain "joint distribution functions"; in this case the elementary probability law gives the probability that the various values of the n-tuple will jointly occupy a set of specified intervals.

d) "Universe" is defined, in logical terms, as the set of all observations which obey a specified elementary probability law. In general, it is assumed that the set is infinite.

[1] For example, the "events" might be the number of deaths in a certain area, or the number of red corpuscles in a certain selection of blood.

e) "Independence of events," which has been defined in the previous chapter: two events, x_1 and x_2, are independent if, when p_1 is the probability that x_1 lies within an interval I_1 and p_2 the probability that x_2 lies within an interval I_2, then the probability of the *joint* event (x_1 in I_1, x_2 in I_2) is p_1 times p_2.

f) "Random sample," which has been defined in the previous chapter as a sample, all of the elements of which are independent events, drawn from the same universe.

g) "Statistic," which is defined as any mathematical function of the observations. For example, if we add the observations and divide by the sample size, we obtain the statistic called the "sample mean," which is usually taken as the best estimate of the unknown universe mean. Again, we may take

$$s^2 = \frac{\sum^n (x - \bar{x})^2}{n - 1}$$

where $\bar{x}$ is the sample mean, n the sample size. This is called the sample variance, and is an estimate of the unknown universe variance.

h) "Degrees of freedom," which depend upon a given statistic for their meaning. In computing certain statistics, we impose algebraically linear restrictions on the observations; i.e., we require the observations to satisfy equations of the form

$$ax_1 + bx_2 + cx_3 + \cdots = K$$

(where a, b, c, etc., do not all equal *zero*). If n is the sample size, and k the number of independent linear restrictions, then the "degrees of freedom" are defined as $n - k$. For example, the sample mean usually has n degrees of freedom; the sample variance has $n - 1$ degrees of freedom (since the sum of the observations must be equal to $n\bar{x}$).

Besides the demand that the probability theory include these basic concepts, certain demands of completeness must be imposed. These demands are very difficult to formulate precisely, even within abstract mathematics, at the present stage of development. However, from the point of view of application, we can state in rather

general terms what we would like the probability theory to do for us. We would like to be able to determine the probability that a certain property of a *statistic* (see definition g) will occur, provided specific hypotheses about the observations are true. If we can make such decisions, we are then able to determine whether the acceptance of an hypothesis is "at variance" with a given set of observations; if the hypothesis makes us infer that the observations we obtained were extremely unusual, then we have a basis for rejecting the hypothesis (if we accept the observations). This matter will be discussed more fully later on under the problem of decision. Put in formal language, we would like to be able to satisfy the following demand:

"Let $O(x_1, x_2, \cdots, x_n)$ be any random sample, with known elementary probability law; let t be any statistic of the sample with degrees of freedom at least 1; then the theory should be able to state the elementary probability law of t."

At the present time even more restricted formulations of this demand cannot be met. At the same time, statisticians have begun to generalize their problem, so that we can answer questions about a sample, even though we do not presuppose an elementary probability law, but rather only presuppose certain of its general characteristics (such as continuity of the distribution function). Such "nonparametric" problems would actually require a much stronger demand for completeness than the above (15).

In lieu of a general definition of completeness, we make a less precise one, pragmatic in character, that the probability theory be rich enough to enable us to make decisions on problems concerning sample statistics in practical cases.

In summary, the probability theory is designed to describe in mathematical terms the probability of certain events or observations, provided we grant specific hypotheses concerning the manner in which the events may occur. These hypotheses, from which we deduce the chances of certain occurrences, usually describe the long-run frequencies with which a random observation will occur within a certain interval in some scale of measurement. The general point of introducing probability theory into the method of science is to enable us to see whether a given hypothesis is or is not "in agree-

ment with" a set of observations. Now, the hypotheses with which we are concerned make reference to a "universe" of elements, and in general such universes are infinite in some precisely defined sense. Hence, a finite set of observations can never lead us absolutely to reject a given hypothesis about such a universe, even if we grant the "truth" of the observations. But we *can* deduce, if the hypothesis is made general enough, and the probability theory is strong enough, the probability that a certain characteristic of the observations will occur. That is, if we let H stand for the hypothesis, and x for a statement about a specific set of observations, we can deduce within the formal probability theory a proposition: "If H is true, then x occurs with probability p." Now we grant that x has occurred, on the basis of granting the specific observations. If p is small, i.e., if x is an unlikely event assuming H, we are inclined to reject H. The method is analogous to *modus tollens:* "If p is true, q is true; but q is false; hence p is false," except that in this case "q is false" never occurs, but rather, "q is improbable." The reader will undoubtedly have several immediate questions to ask: how "improbable" should the consequence be before we reject H? Is this method of rejection general? But these questions had best be postponed until we have examined the other phases of the method.

We now proceed to the second requirement for giving a response to a question:

2. *There must be given a set of observations.* It is to be noted that the basic theory of probability discussed above makes no reference to a *given* set of observations or events. The theory merely describes the probability that certain events will occur, *if* one grants specific hypotheses. The theory is thus concerned with the properties of observations in general, and does not presuppose any specific set of observations to exist. This distinction can perhaps be made clear by stating that we do not talk about the "probability" of a given observation. If we grant that a coin has been observed as "heads" at a certain throw, we do not want to talk about the probability that the occurrence was heads; granting the "validity" of observations, their probability is always one. The probability theory, however, from its general point of view, can talk about the probability that a certain observation, drawn from a well-defined

universe of possible observations, falls within a certain interval, or has some other property. The theory is not concerned with whether an observation has or has not occurred, but only with the question how frequently a certain possibility out of a universe of possibilities will occur. Thus, even though an unbiased coin has been *observed* to fall heads five times in a row, the *probability* that it will fall heads on the next throw is still one-half, for such probability is deduced within a general theory that makes no assumptions about given observations.

The pertinent question we shall want to keep raising throughout this essay is "What are observations?" And the reply to this question, we shall find, is a necessary part of our philosophical reply to all the questions we will raise regarding statistical methods. In general, the methodology we are describing merely assumes that a set of observations has been given which are "pertinent" to the original question asked. No general criteria of pertinence are given, in the sense that we are told how to perform the operations necessary to get an observation. Such questions of an operational sort we must postpone for later discussion.

But statistical theory does impose its own kind of demand upon the observations. As an example, let us suppose that an experimenter wishes to compare two types of thing, say two metals with respect to their tensile strength. How many observations should he make? Should he make the same number of observations on each type? These are questions concerning the adequacy of a set of observations that are at least partially answerable by statistical theory. In the more general case, there may be many types of things under study, all subject to change over the period of observation, and each observation itself an n-tuple of numbers. We want to know the "best" way to make the observations and the number of observations that should be made. The question concerning the way in which the observations should be grouped has been called by R. A. Fisher the question of "experimental design." [1]

In a sense, the problem of the best "design" of an experiment is exactly the problem of the philosophy of science, so that Fisher's

[1] See, for example, (7) and (8). In addition to Fisher's work, there has been extensive work done by others, especially in the field of agriculture.

designation might lead one to suppose that he was formulating replies to the fundamental problems of this essay. But Fisher's meaning of design has nothing to do with the technique of making observations, or the formal presuppositions we bring to bear on an experiment; he presupposes that certain observations can be made, that they are pertinent in general to the question asked, and that the observations obey certain probability laws. He then attempts to solve the statistical problem: how to group the observations so that we obtain the "maximum information" for a given number of observations. "Maximum information" is an ambiguous term, but we may readily give illustrations. In general, when one is comparing two metals, say with respect to hardness, and the readings on each type are subject to the same dispersion, the maximum information is obtained if we make the same number of observations on each type (rather than twice as many observations on one than on the other); this means that we are less likely to make a mistaken judgment as a result of the observations.[1] In more complicated instances, we may have a whole set of questions to which responses must be given, and we want the most informative observations with respect to all the questions. As in the case of the other concepts under discussion, we shall want to discuss "most informative" in a general manner in later chapters.

The solution of the problem of the number of observations that ought to be made in order to respond adequately to a question evidently depends upon the meaning one wishes to assign to "adequacy." If the risk of making a wrong selection is very serious, and ought to be minimized to the greatest feasible extent, then the sample size may have to be very large, especially if one is trying to distinguish two things that are very much alike relative to the error of observation. Recently, the statistician has generalized his method so that, rather than assign a fixed sample size at the beginning, we let the observations themselves decide for us at what point the experiment should stop. This method, now commonly called "sequential analysis" (21) does usually lead to an

[1] For example, one statistic gives "more information" than another (both being based on the same set of observations) if it has a smaller variance, in the technical sense used above. This definition, however, is not completely general.

economy of sample size in the long run, but its present field of application is very restricted; further, to apply sequential analysis, it is usually necessary to be more specific about the meaning of a question than one has to be when a fixed sample is used. The meaning of "more specific" will be discussed below in connection with the next two aspects of the method.

One necessary condition we must impose upon the number of observations: if the method of making an inference from the observations involves a certain statistic which has $n - k$ "degrees of freedom," then we require that the number n of observation be greater than k.

The result of this discussion is that in general the meaning of observation, i.e., the criteria of adequacy in connection with observations, depends upon the meaning of the question asked; so that no question is specific if it leave ambiguous the criteria of adequacy of observation. This point will be made more general as we proceed.

We now consider the third task of one who would precisely formulate a question to be answered by observation:

3. *In terms of the question asked, alternative hypotheses are to be constructed.* This procedure represents the heart of the recently developed techniques, and a rather thorough understanding is necessary before one can grasp the significance of the revolution mathematical statisticians have accomplished in the field of experimental inference.

As a beginning to this discussion, we make use of the example cited in the first chapter. Suppose the original question, then, to be concerned with the similarity or dissimilarity of two types of products. The original formulation of the question is: "Does the special type have a higher mean breaking strength than the regular type?"

This formulation is somewhat ambiguous from the point of view of application. Suppose the (true) mean breaking strength of the regular is 5.25000 ⋯ and the (true) mean breaking strength of the special is 5.250001000 ⋯. Would this information be of any value to the experimenter? If not, then the original question should be reformulated somewhat as follows: "Does the special type have the same mean breaking strength, or a breaking strength that is greater by at least an amount k?" (where k is some figure that "makes a difference" to the questioner).

Now in general we can best make explicit what we mean by a question, by formulating what we take to be the possible answers. In other words, a question remains ambiguous until one can state what the possible answers are like. This means that when we ask a precise question, we are in effect proposing a number of alternative hypotheses about the natural world. In the example given, one form of the question would be:

"Which of the following hypotheses is correct: H_1, The two products have the same (true) mean breaking strength; or H_2, The two products have different mean breaking strengths?"

Another question on the same subject would be:

"Which of the following hypotheses is correct: H_1, The two products have the same mean breaking strength; or H_2, The special product has a mean breaking strength that exceeds the mean breaking strength of the regular product by at least an amount k?"

Now the next problem is to determine how in general we should construct those hypotheses which represent all possible answers to a question. The technique of construction evidently depends upon the remaining step of the method: the selection of one hypothesis and the rejection of the remaining, on the basis of a set of observations. From this point of view, the simple formulation given above is not complete enough, for we are not given any information about the character of the observations. Such information cannot be derived from the probability theory, for the theory is general and makes no assertion holding only for a specific set of observations. Again, the observations themselves are simply number-groups, or else are elementary (singular) propositions, and do not contain any information about the kind of universe from which, or the manner in which, they were drawn. Since we hope to apply the basic probability theory in selecting one hypothesis from among several, we must include information about the manner in which the observations are subsumed under the probability theory.

Let us consider an example in which the hypotheses have this general property. Consider first the following set of proposals:

1. The observations on each product are those given in Chapter I, suitably divided into two samples, R, the regular, and S, the special.

2. The observations in R constitute a random sample, and the observations in S constitute a random sample.

3. The random sample R comes from some normal universe, and similarly the random sample S comes from some normal universe.

4. The mean of the R-universe is the same as the mean of the S-universe.

5. The variance of the R-universe is the same as the variance of the S-universe.

Then *one* question about the products would be formulated in terms of the following hypotheses:

H_1: accepts 1, 2, 3, 4, and 5;
H_2: accepts 1, 2, 3, and 5, but rejects 4.

If we add to the list

4*: The mean of the S-universe exceeds the mean of the R-universe by at least an amount k ($k > 0$), then *another* question would be [1]

H_1: accepts 1, 2, 3, 4, and 5;
H_2: accepts 1, 2, 3, 4*, and 5.

If the question is put in more general terms, so that one asks whether there is *some* difference between the regular product and the special, we would have

H_1: accepts 1, 2, 3, 4, and 5;
H_2: accepts 1, 2, 3, but rejects either 4 or 5 (or both).

We now proceed to a general formulation of the technique of constructing alternatives. It must be understood that this technique is relative to the present methodology in mathematical statistics, and that developments might readily demand revisions.

In order to construct an adequate set of alternatives, i.e., in order adequately to specify the meaning of a question, the following demands are to be met:

a) *The hypotheses are to include enough information about the universe from which the sample was drawn and the manner in which it was drawn, so that a precise method for selecting one of the alternatives*

[1] In this case H_1 and H_2 are not exhaustive of the possibilities, and technically a third hypothesis would have to be included.

can be set up, on the basis of the fundamental probability theory and the observations.

The adequate explanation of a) must wait upon the explanation of the last stage of the general method, the selection of an alternative; the demand is formulated here to show how intimately connected are these two aspects of experimental inference. Recent work in mathematical statistics has shown that the information required by a) may be put in a very simple form; for example, the following question could be formulated apropos of the comparison of the two products:

1 and 2 as on pp. 24 and 25.

3. The random sample R comes from a universe having a continuous distribution function (in the mathematical sense of "continuous");[1] likewise for the S-sample.

4. The R-universe and the S-universe are the same (i.e., obey the same elementary probability law).

Then H_1 accepts 1, 2, 3, and 4, H_2 accepts 1, 2, and 3, but rejects 4. Such formulations of questions, in that they do not compare universes by parameters (e.g., the mean, or variance), are called "nonparametric."

Besides this very general requirement in the formulation of the alternatives, there are a number of other requirements of a logical or formal nature.

b) *The propositions in each alternative should be expressed in the language of the basic probability theory.*

c) *Each alternative should be expressed in the form of a logical conjunction of propositions with the following property: no member of the conjunction is deducible from any combination of the others, or from the others in conjunction with the postulates of the probability theory and the observations.*

d) *Each alternative must be consistent, both in itself and with respect to the probability theory and the observations.*

The demands in b), c), and d) in effect are requirements that each alternative be a "formal system" of independent and consistent

[1] More generally, one can replace this condition by the condition that the distribution function is continuous "almost everywhere," i.e., the set of points of discontinuity is at most denumerably infinite (15).

postulates; the basic language of each of these formal systems is specified to be that of the probability theory. The consistency and independence are all made relative to the basic probability theory, as well; in effect, we must not be able to demonstrate either the truth or falsity of any alternative within the formal probability theory.

In the sequel, we shall call the propositions of the original formulation of the alternatives the "fundamental propositions," to distinguish them from the theorems that may be deduced within a given alternative.

We now proceed to a discussion of the relationship that must hold *between* alternatives:

e) *The alternatives should exhaust all possible answers to the question.*

This requirement simply asserts that the alternatives should completely define the question asked. What "all possible" means in this connection will evidently depend upon the meaning of the question, i.e., upon the intentions of the questioner.

f) *No two alternatives when conjoined should be consistent.*

e) and f) therefore demand that the alternatives form an exhaustive and mutually exclusive set of opinions on the question asked.

For the sake of convenience in analyzing the alternatives we make the further demand:

g) *A fundamental proposition of one alternative must be the logical equivalent or else the logical contrary or contradictory of a fundamental proposition in any other alternative.*

This means in effect that we can "pair" the propositions of any two alternatives, and that each pair are either equivalent, or else are contraries or contradictories. This demand is one of convenience.

In agreement with philosophical usage, we can regard those fundamental propositions which are the same in all alternatives (i.e., the logical equivalents of which appear in each alternative) as the *presuppositions* of the method. The remaining propositions are regarded as the basis of *inquiry*, in that they represent the heart of the question asked. In the illustration given above, in the first formulation of the question, assertions 1, 2, 3, and 5 are presuppositions, assertion 4 is the basis of inquiry.

We make, as a final requirement, the "obvious" demand:

h) *At least two alternatives must be given.*

The conditions we impose on the alternatives constitute a definition of a "precisely formulated" question. Such precision, however, is relative to statistical theory at this stage; an *absolutely* precise question would have to show the general meaning of "observation," "random sample," and so on, i.e., would have to show how these concepts gain meaning in the activities of the experimenter. This is the problem to be discussed in the subsequent chapters.

The final, and perhaps most difficult, task of the method is concerned with the method of selecting one of the alternatives:

4. *A method must be formulated for "selecting" one of the alternatives on the basis of the probability theory and the observations.*

The conditions we have placed on the alternatives make it impossible to select one over the others on the basis of logical or formal deduction alone. Evidently we require some technique of selection that differs radically from that used by the formal scientist.

Our first task will be to state rather "obvious" demands to be imposed on any suitable method of selection.

Suppose an experimenter asserts that, regardless of what observations he obtains, he intends to select one of the alternatives. His procedure would certainly constitute a well-formulated method of selection, but common sense suggests that such a procedure would be futile in the long run. In order that the demand for observations have meaning, we must impose the following condition on the method of selection:

a) *The information contained in the observations is to be made a necessary condition for the selection of the alternatives.*

Again, common sense demands that we treat all the alternatives "fairly," i.e., "with an open mind." This suggestion evidently requires a translation into the language of mathematical probability to be of any value to the method we are considering. The translation suggested by statistical theory is the following: a method of selection is said to be *unbiased* if the risk (i.e., probability) of rejecting any given alternative is at a minimum when the alternative is (in fact) true. For example, suppose the question is formulated in terms of k exhaustive and exclusive hypotheses H_1, H_2, $\cdots$,

H_k. We now deduce, for a given method of selection, the chance of rejecting H_1 when it is really true; call this chance α_1. We also deduce the chance of rejecting H_1 when one of the other hypotheses $H_2, H_3, \cdots, H_k$ is true; call these chances $\beta_{12}, \beta_{13}, \cdots, \beta_{1k}$. Then we require of an *unbiased* method of selection that α_1 shall not exceed any of the quantities β_{1i} $(i = 2, 3, \cdots, k)$.[1]

To illustrate, suppose we ask whether a coin is biased or not. Suppose, further, to make the question simple, that we know a priori that the coin is one of three types: unbiased, biased to heads in the ratio 3 to 4, biased to tails in the ratio 3 to 4. The question is then dependent upon the following assertions:

1. Any set of throws constitutes a random sample.

2. The probability of exactly k heads and $(n - k)$ tails in n random tosses is given by[2]

$$\frac{n!}{k!(n-k)!}\left(\tfrac{1}{2}\right)^n.$$

2′. The probability of exactly k heads and $(n - k)$ tails in n random tosses is given by

$$\frac{n!}{k!(n-k)!}\left(\tfrac{3}{4}\right)^k\left(\tfrac{1}{4}\right)^{n-k}.$$

2″. The probability of exactly k heads and $(n - k)$ tails in n random tosses is given by

$$\frac{n!}{k!(n-k)!}\left(\tfrac{1}{4}\right)^k\left(\tfrac{3}{4}\right)^{n-k}.$$

Then H_1 accepts 1 and 2, H_2 accepts 1 and 2′, H_3 accepts 1 and 2″. Suppose now, on the basis of three observations, that we formulate the following method of selection:

Select H_1 if and only if the number of heads is 2. Select H_2 if and only if 3 heads appear. Select H_3 if and only if the number of heads is 0 or 1.

Such a method of selection is biased, according to the definition

[1] If the number of hypotheses is infinite, then β_{1i} becomes a function of one or more variables over an infinite set. For example, in Student's test of means, β is a (monotonic increasing) function of the (true) absolute difference in means of two normal universes.

[2] This elementary probability law defines a "binomial distribution," having but one parameter, p, which is the chance that a single event will occur in a series of independent tries.

above. The probability of rejecting H_1 when it is actually true can be shown to be $\frac{5}{8} = \alpha_1$; on the other hand, the probability of rejecting H_1 when H_2 is actually true can be shown to be $\frac{37}{64} = \beta_{12}$. Hence $\alpha_1 > \beta_{12}$ and, according to this method of selection, H_1 has a better chance of being selected when it is false.

An example of an unbiased method relative to H_1 (though not necessarily the "best") would be:

Select H_1 if and only if the number of heads is 1 or 2. Select H_2 if and only if the number of heads is 3. Select H_3 if and only if the number of heads is 0.

If we grant that a sound method of selection should be unbiased, then we should impose two more conditions:

b) *The method of selection and the alternatives should be so formulated that one is able to deduce for any given alternative the probability of its rejection when it is true, on the basis of a specified set of observations, and the probability of its rejection when one of the alternatives is true.*

The probability of rejecting an alternative when it is true is called the "type I" error associated with the alternative; the "type II" error is the probability of accepting the alternative when it is false, and is evidently $1 - \beta_{1i}$, its magnitude depending upon which of the alternatives is true.

c) *The method of selection should be unbiased.*

It will be noted that b) really imposes conditions on all four aspects of the methodology: the probability theory, the observations, the formulation of alternatives, and the method of selection. It is further to be noted that condition b) has yet to be satisfied in a very wide range of applications of statistical techniques. However, even if b) is not satisfied, we can establish condition c) provided we establish a weaker form of b): that whatever may be the probability of rejection, it is possible to show that it will be at a minimum when the alternative is true.

There are some cases in which all the experimenter need do is specify the method of selection, the set of alternatives, and *one* chance of error for a specified alternative;[1] then the chances of error

[1] This alternative is usually called the "null" hypothesis in statistical literature.

associated with the remaining alternatives will be deducible. Such is the case in most of the "classical" methods of selection developed by Student, K. Pearson, R. A. Fisher, and others. When, as in the case of sequential analysis, the sample size is not fixed a priori, it is usually necessary to fix at least two chances of error, from which the character of the remaining chances may be deduced.

In addition to the requirement for a lack of bias, other demands occur to us that should be imposed upon a "sound" method of selection. There is on the one hand the formalist's general demand that all the methods of selection should spring from one central concept, so that the various methods one employs are not chosen in an arbitrary or haphazard manner. There is, on the other side, the pragmatic desire to have the method make the most "useful" selection.

The demand of generality of method is analogous to the well-known *decision problem* in formal theory. The decision problem consists in showing that for any well-defined proposition x of a formal system, either the assertion "x is provable" or else the assertion "non-x is provable" is true. If we can show this, we assert that the formal system is complete: there are no insolvable problems. If we cannot show this, then the formal system is not adequate for the answering of all meaningful questions, according to its own criterion of meaning.

Suppose we rephrase the formal decision problem so that its analogue to our present discussion can be made clearer. We say that a well-designed formal theory must satisfy at least the following demands:

1. It must show how meaningful "propositions" are constructed out of a basic set of terms, operations, and relations. A proposition is defined as a form to which we may assign either the logical value "true" or the logical value "false." Propositions are thus distinguished from "propositional functions" (such as $x + 2 = 7$) which are ambiguously true or false, depending upon the specific meaning of certain variables occurring in them. For example, an adequately formulated arithmetic should specify how "meaningful" propositions can be constructed out of the numbers, the operations "plus," "times," etc., and the relations "equals," "less than," etc.

2. It must show how theorems are constructed out of a basic set of axioms. Such "construction" is usually called a process of "deduction," but viewed abstractly the construction simply shows how the axioms are to be transformed into other assertions to which we assign the logical value "true." Further, by the simple operation of logical contradiction on the true propositions, we can construct propositions to which we assign the value "false."

On the basis of 1 and 2, we may always propose the following question: Are there propositions constructable by the criteria of 1 for which we can neither assign the value "true" nor the value "false" by the criteria of 2? If so, then the formal system is incomplete, and its decision problem is not solvable.[1]

Analogously, in the theory of statistical inference, we demand:

1. A method of formulating alternative hypotheses relative to any question, each hypothesis being expressible in the language of probability theory, and we require a corresponding set of "observations."

2. A method of selecting one hypothesis from a well-formulated set of hypotheses and a set of observations.

We may then ask whether a method that satisfies 2 will be applicable to every possible set of hypotheses formulated by the criteria of 1. On grounds of generality, we would like to require that

d) *The method of selection should be adequate for a complete solution of the decision problem.*

To date, the decision problem of statistical inference has not been solved, provided we impose the condition that the method of selection be unbiased.[2] It has turned out that the commonly used techniques of selection become extremely complicated in certain instances, and involve the mathematician in as yet unsolved problems within his formal domain. It is interesting to note that the desire for *one* basic method has been challenged by the conflicting desire for simplified techniques. Even though we might find one basic method, we

[1] Technically, there are several formulations of the decision problem which have been made but which, for the present purpose, need not be distinguished.

[2] For example, if this condition were not imposed, we might offer as a general method something trivial of the following sort: after the hypotheses have been ordered by a certain technique, select the first one in every case. The ordering would have to be based upon the observations, in order to satisfy a) on p. 15.

would probably find that its application was too complicated in many cases to warrant its universal use.[1] It appears that some of the modern computational devices may alleviate to some extent this demand for simplification.

The pragmatic desire for a simplified method of selection raises the general problem of what we want the method to do for us. Can we enunciate definite criteria for good and bad methods? Must we not admit that these criteria of value are relative to the special purposes of the investigator?

These are general questions about the methodology which will be postponed for later discussion. It is interesting to note, however, that in certain cases mathematical statisticians apparently have been able to formulate *sufficient* conditions for a best method of selection; that is, if the conditions are satisfied, then the method is best, though the converse cannot be made to hold. An example of such a best method is the Student test discussed earlier. This method has the following strong property: let α_1 be the long-run chance of rejecting a certain alternative H_1 when it is true; then the chance of accepting H_1 when it is false is at a minimum for this method as opposed to all other methods, regardless of the manner in which H_1 is false, i.e., regardless of which of the other alternatives is actually true. Such methods are called "uniformly most powerful," since their "power" (as measured inversely by the chance of acceptance when false), is maximum compared to that of any other method no matter what may actually be the truth. Uniformly most powerful tests do not always exist, so that to make them a necessary condition for a method of selection is to deny in a very disastrous way the demand for generality.

It has turned out to be a very difficult task to find criteria for selection-methods that will have the force of uniformly most powerful tests. Reference 2, at the end of the chapter, reviews some of the formal attempts to find substitute conditions.

In References (1) and (19), certain more recent general definitions are discussed. Wald suggests that we associate with each hypothesis a certain "weight function." In simplified terms, this function

[1] For example, the method employing the so-called "likelihood ratio" (13) often involves equations which have to be solved by lengthy approximation techniques.

expresses the loss we may expect to incur as a result of our method of selection. That is, if a given hypothesis is true, there will be a certain loss associated with accepting any one of the alternatives. This "loss" will presumably be zero, if we accept the true alternative, and will usually increase as we accept alternatives "farther" from the truth.

As an example which is obviously to the point, suppose we wish to find out whether a die is biased. The alternative hypotheses are usually infinite in number, and are of the form:

> H_p: The throws are random, and the (true) probability of a six (or any other number) appearing is p; where p can take on any value from 0 to 1.

Suppose now that the die is, as a matter of fact, unbiased, and we proceed to bet on the basis of a certain hypothesis H_p. If we accept the correct hypothesis, the "loss" incurred from our decision will in the long run be zero; if, on the other hand, we accept the hypothesis that the true probability is $\frac{1}{7}$, then the loss will not vanish, but will be less than the corresponding loss if the true probability is, say, $\frac{1}{2}$ or $\frac{3}{4}$.

We can now say that the loss we incur in making a given selection is a function of the loss involved in a given choice, and the probability that such a choice will be made under a given method of selection.

In terms of this generalized meaning of loss, Wald and Brookner suggest certain criteria for best methods of selection (1) and (19). Both admit that other "reasonable" definitions are possible, and this attitude, together with the problem of assigning values to the "loss," raises the question which will push us on to the discussion of basic issues: can we expect to solve the problem of the best method on formal grounds alone? The answer clearly seems to be negative, and yet a negative response raises all kinds of issues of an awkward sort.

In sum, the survey of this chapter, designed to describe briefly the most exact methods of inference modern science has been able to develop, will raise for us, in the next chapter, problems that indicate how very incomplete these methods actually are. Our

task will then be one of suggesting how we can preserve the value of these modern advances in the light of a more general description of scientific methodology.

Note: Though the next chapter will raise the problem of the generality of the methods discussed here, there is nevertheless one aspect of generality that should be mentioned in connection with the present treatment. Within modern statistical theory, one usually makes a distinction between the theory of tests of hypotheses and "estimation-theory." Briefly, the former considers methods of selecting one from a set of alternative hypotheses, while the latter considers methods for deciding the "best" value for a certain unknown measurable quantity, or, more generally, for deciding the best interval within which a quantity is thought to lie. For practical purposes, it is certainly advisable to keep these two aspects separate, but from the general point of view of the present treatment such separation evidently has serious disadvantages. It is the purpose of this note to show how the theory of estimation is subsumed under the methodology developed in this chapter; the material has been incorporated in a note since the subject matter is somewhat more specialized than the general treatment of the essay.

The problem of estimation actually occurs in several distinct phases in probability literature, as follows:

a) We may wish to find a single number which in some sense represents the best guess as to the true (but unknown) value of a measured object;

b) We may wish to find that interval which in some sense is most likely to include the true (but unknown) value of a measured object (the so-called problem of Confidence Intervals);

c) We may wish to find that interval within which the means of future sets of observations are in some sense very likely to lie (the so-called problem of Tolerance Intervals).

In the classical treatment of errors, b) and c) are very often confused (e.g., in the representation of the mean plus or minus the probable error). Further, the value obtained to satisfy a) is sometimes erroneously called "the most probable value," as though the true value were a random variate.

Now if the unknown quantity is regarded as one of an infinite set of numbers (denumerable or not), then in general the problem proposed under a) is meaningless. The chance of our selecting the correct value from such an infinite set [1] on the basis of a finite number of observations is zero. Actually, when we make single estimates of the type considered under a), the estimates are used as a basis for selecting some hypothesis, or solving problems under b) and c). Such estimates are "incomplete" descriptions, and do not represent any decision on the part of the experimenter concerning the natural world. It is true that the literature refers to these estimates as "maximum likelihood estimates," "least squares estimates," etc., as though they possessed some value in themselves, but a careful examination of context will show that such estimates are "good" only in so far as they enable us to solve decision problems of the sort already discussed in connection with hypotheses.

If the true value can only be one of a finite set of values, then the problem is a special case under b).

Problem b) has led to some considerable confusion in statistical literature, chiefly because the exact formulation has been difficult to give. A naive formulation would be: "What is the interval which very probably contains the true value?" Such a formulation is naive on several counts: what does "very probably" mean? how is the size of the interval to be restricted? But a more basic objection is that the true value is not a random variate, that it is a unique element among the real numbers, and that the probability of its lying in *any* interval is therefore either exactly one or exactly zero.

If the probability concept is to be used in describing the interval we wish to choose, it must be used in connection with the only things that can be random, the observations. Thus viewed, the naive formulation takes on the more sophisticated form: how shall we select an interval which will best explain the kind of observations we have obtained?

The problem of b), i.e., the problem of confidence intervals, has now been reduced to a problem quite similar to that discussed in this chapter. It is certainly useful within the formal theory of

[1] The set should have the characteristic that all the elements have non-zero weight, i.e., that no possibility is actually excluded by the observations

statistics to keep the problem of interval estimation and the problem of hypothesis testing separate; but from the point of view of the general discussion of this essay, the two problems are methodologically alike, in that they raise exactly the same general problems discussed in the next chapter. Similar remarks apply to problem c), the problem of tolerance intervals that are so widely used in industrial quality control.

REFERENCES

1. Brookner, R. J., "Choice of One among Several Statistical Hypotheses," *Annals of Mathematical Statistics* (1945).
2. Camp, B., "Some Recent Advances in Mathematical Statistics, I," *Annals of Mathematical Statistics* (1943).
3. Churchman, C. W., "Logic of Statistical Tests," *Bulletin of the Institute of Experimental Method* (U. of Pa.), (1947).
4. Cramér, H., *Mathematical Methods of Statistics*, Princeton Univ. Press (1946).
5. Eisenhart, C., "The Assumptions underlying the Analysis of Variance," *Biometrics* (1947).
6. Ferris, C. D., Grubbs, F. E., and Weaver, C. L., "Operating Characteristics for the Common Statistical Tests of Significance," *Annals of Mathematical Statistics* (1946).
7. Fisher, R. A., *Design of Experiments*, Oliver and Boyd (1937).
8. Fisher, R. A., *Statistical Methods for Research Workers*, Oliver and Boyd (1937).
9. Freeman, H., *Industrial Statistics*, John Wiley (1942).
10. Hoel, P., *Introduction to Mathematical Statistics*, John Wiley (1947).
11. Kendall, M., *Advanced Theory of Statistics*, Lippincott (1944).
12. *Methods of Making Experimental Inferences*, Frankford Arsenal, Phila. (1946).
13. Neyman, J., and Pearson, E., "On the Problem of the Most Efficient Tests of Statistical Hypotheses," *Philosophical Transactions, Royal Society*, Series A (1933).
14. Neyman, J., and Pearson, E., "Contributions to the Theory of Testing Statistical Hypotheses," *Statistical Research Memoirs*, Vol. 1, (1936).
15. Scheffe, H., "Statistical Inference in the Non-Parametric Case," *Annals of Mathematical Statistics* (1943).

16. Snedecor, G. W., *Statistical Methods* (4th Ed.), The Collegiate Press (1946).
17. Student, "On the Probable Error of the Mean," *Biometrika* (1908).
18. Tippett, L. H. C., *The Methods of Statistics*, London, Williams and Norgate (1937).
19. Wald, A., *On the Principles of Statistical Inference*, Notre Dame Univ. (1942).
20. Wald, A., "Asymtotically Most Powerful Tests of Statistical Hypotheses," *Annals of Mathematical Statistics* (1941).
21. Wald, A., "Sequential Analysis of Statistical Hypotheses," *Annals of Mathematical Statistics* (1945).
22. Wilks, S. S., *Mathematical Statistics*, Princeton University Press (1943).

Chapter III Problems of Method

The discussion of the preceding chapter on the methodology of experimental inference brought to consciousness a number of problems which were postponed for later treatment. Following the form of the outline, we can enumerate these problems:

a) What justification do we give for the use of one formulation of probability theory rather than another? Does it matter, indeed, which formulation we choose? Are not probability theories to be regarded as modes of language-expression, so that one formulation can always be translated into another? Or, instead, must we regard a given theory of probability as presenting a partial account of the natural world, an account of the manner in which aggregates of certain types of events occur, and therefore an account that differs in essential respects from the account presented by an alternative formulation?

b) According to the methodology described above, one is to select a set of observations taken to be pertinent to the problem. By what criteria do we judge such pertinence? Is it intuitively self-evident that a given observation can be used as evidence in making a decision? Or instead, do we not have to formulate the conditions for the adequacy of observations, conditions that are "operationally defined" in terms of specific behavior patterns?

c) In what manner are we to construct the alternative hypotheses relative to a given question? Note further that these alternatives are all expressed in the language of the probability theory, but the original question is rarely phrased in such language. For example, the engineer wishes to know whether one type of steel has a lower tensile strength than another. The statistician inquires whether

two sets of observations come from the same universe. Some translation is required from the engineer's language to the language of the statistician; how is such a translation to be accomplished?

Further, in formulating the alternatives, we make certain "presuppositions," i.e., certain statements are accepted as true in every alternative. What justification do we have for making such presuppositions? In particular, in all applications of the method described above, it is necessary to assume randomness for certain sets of observations. If randomness cannot be presupposed, then there can be no application of the basic probability theory, for such theory is necessarily concerned with the properties of random variates. What does randomness mean with respect to our experimental method? Are we forced to assert that the world of nature exhibits a random character, which we "find" by our observations? Or is randomness a purely mathematical construct we impose on the natural world, the construct itself having no reality outside the scientist's method? Or is randomness to be used as a basis for deciding the adequacy of observations, so that no observation can be regarded as pertinent to a problem unless it is made by a specific set of operations designed to give a random result?

d) Finally, by what principles do we choose the method of selection of one alternative over all the others? Do we formulate what are intuitively reasonable criteria in terms of risks? Or must we attempt to evaluate a given method with respect to our particular purposes associated with a specific problem? Or must we first formulate the general criteria of value, and then specify how these ethical principles are to be applied in determining a best method of selection?

These, and their like, are the kinds of questions that occur to the methodologist in attempting to apply the techniques of mathematical statistics. They are problems whose solution is not to be found within statistical theory, and yet they are also problems whose solution is necessary for the application of such theory. For the moment, we shall call such problems "philosophical" in character, not meaning thereby that they involve any metaphysical concepts, but only that they are in some sense general with respect to the individual sciences.

The attempt will be made in this essay, not so much to give replies to the specific questions raised above, but rather to formulate a general version of scientific method within which the specific problems can adequately be studied.

For this reason, instead of proceeding immediately to consider problems such as the meaning of randomness, we shall propose a more general problem. This problem appears to be basic to the philosophical study of method, and the various answers that have been given to it we shall take to characterize the philosophical opinions on the subject of scientific method.

The problem I have in mind is one that must long before have occurred to the reader: Is the methodology of inference described in the previous chapter a general one? That is, must such a method be employed in the answering of *all* questions the mind can raise?

This problem is basic to a discussion of the specific problems we have raised in connection with the four aspects of statistical method. If we can answer some questions by techniques different from those given in Chapter II, then we should incorporate these techniques into the general study of inference, before developing any special aspect of the method. Perhaps, for example, the basic probability theory is "given" to the experimenter by a rational intuition of its validity. In such case, we would have to incorporate the activity of such intuition into the general method, and such incorporation might well enable us to answer questions about the other three phases.

If, on the other hand, we should argue that the method of Chapter II is general, then we propose for study the very difficult problem of how science can operate and progress under such conditions. The replies we then make to the special questions would have a quite different character from the one just suggested.

Now, in considering the problem of the generality of statistical methods, we should take at least one lesson from the discussion of Chapter II. The problem is still in a vague form, and should be restated in order to conduct an adequate study. Further, such a restatement should enable us to set up alternative answers which in some sense exhaust the possibilities.

The restatement we shall make will depend upon selecting the

one aspect of statistical method that is necessary for its application and appears to stand in the road of its general application: must one make presuppositions in the answering of any question? The term "presupposition" is used here, not in the vague sense of the a priori in metaphysical literature, but in the technical sense used above: a presupposition is a proposition which (implicitly or explicitly) is asserted in all the acceptable alternative responses to a question.

This aspect of statistical method has been made the basis for the subsequent philosophical study because it has been the one most emphasized in the history of philosophy. And this emphasis is not in the least surprising for, as we shall see, the rôle of presupposition is in a very real sense basic to our understanding of science.

The problem of presupposition actually involves a correlative problem: whether or not *any* answers can be given to questions. These two aspects of methodology will be the subject of the next chapters.

The manner in which we shall conduct this study of the two basic problems, what can be answered and what must be presupposed in order to give answers, demands some consideration. A natural tendency would be to attempt an examination of what science *now* does. Such an attempt would involve an analysis of present-day practices, and a generalization of these with the purpose of determining the basic principles of experimental methods.

But an attempt at analysis of this sort is doomed to failure on several counts. In the first place, we require some basis for the proposed examination of present practices, and such a basis surely depends upon the manner in which such practices have developed, i.e., upon an historical study of methodology. But in the second place, the philosophical purposes of this essay are evidently not satisfied by an empirical inquiry into the behavior of experimental scientists within one culture. Such an inquiry would properly belong to anthropology, and its bearing on the general problem of methodology would be only indirect.

The philosophical problem we have raised implies that we should consider method from the point of view of an *ideal:* we wish to examine what scientists *should* do if they are to provide adequate

answers to questions. We are only indirectly interested, therefore, in what they actually *do*, for their present behavior may not (often is not) best designed to serve this basic end.

This observation suggests a method of study of the problem at hand. We have borrowed from Chapter II the demand that an exhaustive set of alternatives be given; we may further borrow the demand that a method be devised for selecting one alternative over all others. We could hardly hope, in so general a study as this one, to reduce the method to the terms of statistical theory. But the discussion above suggests how the alternatives may each be used to serve the end of selecting the most plausible answer. If the study of science's method must be based upon an historical study of theories, then perhaps we can find within history a pattern of progress in the understanding of methodology. If such a pattern can be found, it should serve to indicate how certain of the alternatives failed adequately to account for scientific method, and how one solution is in some sense forced upon us. Our next task is therefore an attempt to make such an historical analysis of the meaning of method.

Chapter IV The Dialectic of Modern Philosophy

It has become an all too common saying that philosophy is a field of research in which it is impossible to detect signs of progress. In what sense, the challenge would have it, can we assert that our present-day problems and our present-day answers are "superior" to those of Plato or to those of Oriental thinking? And the problem, obvious as it is in its formulation, is no easy one to solve. One might assert that it was the central problem of Hegel's philosophy to formulate a criterion of philosophical progress, and that the consciousness of the problem in modern philosophy did not really occur until the post-Kantian period.

The dissatisfaction subsequent thinkers have found in the Hegelian dialectic as a solution to the problem is grounded, I think, on an implicit dissatisfaction with the applications Hegel made of it. Those of us who do not feel the import of the classifications of the *Phenomenology* are apt to overlook the point that these were examples of a method, and that the validity of the method does not rest on the truth or importance of the examples.

The discussion of the previous chapters, together with the Hegelian concept of dialectic, are now to be used to provide us with a technique for studying such general problems as those that philosophy proposes.

Whatever may be the problem we propose, we have agreed that at least this much must be satisfied: we should be able to exhibit in some form what we take to be *all* the possible alternative opinions. Now the criteria of completeness in this respect are to be found within the science of logic; that is, logic should enable us to

decide whether or not a given set of alternatives exhaust the possibilities.

But the dichotomies proposed by the logician would be empty ones indeed if significant content were not supplied by some other source. In other words, the actual formulation of alternative opinions concerning a problem will demand more than the logician alone can supply. The meaning of the problem itself does not depend upon the logician's devices. The situation is the same with respect to the techniques discussed in Chapter II. We can certainly construct alternatives with respect to any question asked, but the significance of a given set of alternatives cannot be supplied by statistical methods alone. The naive assumption that statistics can play such a game has been largely responsible for the bad name the science has acquired in many fields.

What does give significance to a logical classification of opinions? One may equally well ask what gives significance to *any* problem? We shall here introduce a presupposition of our method of inquiry into opinions, namely, that the significance of a problem depends upon its persistence in history or, in this case, in philosophical literature.

We require, then, another criterion of adequate classification which one might call the "historical." We shall take this criterion to have been satisfied if the material for the classification of opinions is derived from the history of the problem; that is, we shall assume that those divisions roughly laid down by the history of past opinions are the fruitful ones. We may call the classification satisfying the logical and historical criteria a logico-historical method, as does Singer (3). It seems necessary to point out, however, that as difficult as it is to make sure that one's classification is logically rigorous, it is even more difficult at present to find a precise method for guaranteeing historical importance. In offering the logico-historical method as a means of analysis of a problem, we are at best suggesting the direction that such analyses should take. We must rely at present almost entirely on the historian's judgment in these matters, until a science of history shall have been developed that permits a more rigorous selection of categories on the basis of the past.

Similar remarks apply to the method for "selecting" or believing in one alternative opinion rather than any of the others. In this respect, we are really no worse off than we were with respect to the procedures discussed in Chapter II, for there are still to be developed even relatively precise criteria for evaluating methods of selecting alternatives. Our only course is to propose a method which can be made more precise as the science of history develops. This is the dialectical method to which we have already briefly referred.

The dialectical method attempts to find an historical order in which the various possible opinions on a given question have appeared, and to determine the basic reasons for the changes in opinions, i.e., the underlying principles of such change. The Hegelian application of the dialectical method is of course totally inadequate for our purposes, since its basic principles are couched in terms difficult to define within the framework of scientific method. But this obscurity in Hegel's writings should hardly prevent our applying a general form of his technique, if it prove useful. Its usefulness will be found in the fact that by means of the dialectic we can discover reasons for abandoning certain opinions, by showing how, within a general historical trend, the opinions proved inadequate to their purpose. The "method of selection" we shall use is therefore based upon the general trends we find dialectically revealed in history; the rejection of opinions is founded on a general historical survey of a problem.

One note of caution should be emphasized, since the present treatment differs so radically from the usual historical studies of philosophy. We are *not* primarily interested in a textual analysis of philosophical literature. We are interested, nonetheless, in what certain schools of philosophy, and their representatives, were trying to say. Now how does one judge what the intentions of an individual were in writing a given text? Our contention is that it is incorrect to suppose the evidence of the text *alone* to be sufficient and consistent evidence of the writer's intentions. In the first place, no man ever wrote a text in which all sentences were directed towards the same ends; indeed, it is scarcely likely that anyone ever wrote a text which was free of conflicting viewpoints. Such "automatic"

writing implies the absence of unconscious motivations that is inconsistent with our present views concerning minds and their operation.

For example, it is an all too common failing on the part of the commentaries on Kant's first *Critique* to attempt to "explain" confusing passages by references to other sections of the *Critique* that are equally confusing. From the point of view we are developing here, an attempt to explain the "Deduction of the Categories," for example, on such a textual basis is an attempt that is doomed to failure.

We take a more general analysis of a writer's intentions to depend upon just such a logico-historical study as we have proposed above. The meaning of a given philosopher's opinion on a certain question will depend upon the meanings we ascribe to other writers, past, present, and future. That is, we must set up alternative possibilities as to his meaning, and select one of these alternatives on the basis of both the evidence of the text, and the manner in which opinions on the question at issue have occurred in history.

To review the proposed method of historical analysis:

1. With respect to any question a set of alternative opinions are to be constructed, the construction to agree with the formal requirements of a *logico-historical* method.

2. The alternative opinions are to be so constructed that on the basis of historical evidence an ordering of such opinions in time can be effectuated.

3. The basis of selection of an opinion will depend upon the inadequacy of the remaining alternatives as displayed in their historical context (2 and 3 form the *dialectical* method).

The discussion of the previous chapters has suggested the problem that we wish to treat by such a method. For this discussion has raised for us the problem of *how* questions are answered in general, and this problem of the "how" implies the correlative problem of *what* questions can be answered.

We therefore take the fundamental problems of the philosophy of science to be: "What can be known?" and "How can it be known?" If one feels that these two phases are not the only ones to be considered under the fundamental problem, he need only take

the present classification to be concerned with part of the picture; in other words, our object here is not so much to defend a definition of philosophy as to expound the meaning of two of its problems: the "what" and the "how" of knowledge.

Before proceeding to the classification, we must explain as well as we can the meaning of the "what" and the "how" of our problem, at least to the extent of providing a language within which we can express the opinions. When we ask "what" can the mind know, we evidently do not require a list of propositions; rather, our formalist wants a set of classes, so that, for example, a given opinion can be expressed as "some, or all, propositions of this class can be known." What we may call our "historical intuition" of the manner in which the past has studied this problem suggests a twofold classification of types of propositions into the particular and the general, or, as we shall prefer, into facts and laws. To define, then, we say that a question of *fact* is a question whose answer is single-valued; for example, a question whose answer would be of the form "A is k units long," where k is a uniquely defined real number.

Our example of a question of fact has intentionally been quantitative, since it appears that only through the use of quantity can one express the uniqueness necessary for the single-valued property of factual questions. That is, questions of the sort: "Does Jones have a pain?" or "Do I see the color red?" are ill defined for experimental science and their correct defining in quantitative terms would enable one to determine whether or not they were factual questions.

We wish the concept of *law* to represent the logical negative of the concept of fact, so that our classification will be exhaustive. We also wish to make the concept of law correspond at least roughly to the common notion of a "general proposition." To avoid paradoxical results, we must also include in the class of facts all finite conjunctions, finite disjunctions, and negations of and implications between factual statements: we do not want to say, for example, that the proposition "A is k units long and B is l units long" is a law. Thus all propositions constructible out of factual propositions by (finite sets of) logical connectives are themselves facts. We may then say that the class of questions of law is the negative of the

class of questions of fact within the universe of meaningful questions.

When the formalist asks, then, "what" questions can be answered, he is really posing two questions: "Do answers to (some) questions of fact exist?" and "Do answers to (some) questions of law exist?" But he to whom such questions are posed will require some elucidation of the concept of "answer" before he can express an opinion on these matters. And this, of all the basic terms of the problem of philosophy, seems most difficult to define. When we shall say a question is answered, we shall mean that a true statement has been made which satisfies the demands of the problem, so that the concept of "answer" entails the concept of truth, and he who would define the one must take on all the burden history has felt in the defining of the other.

However, we can at least phrase the meaning of "answer" in the language already adopted in Chapter II. We have there suggested that the criterion of selecting one alternative response to a question depends upon the "risks" and "losses" involved. On this basis, we can propose the first presupposition of our inquiry, namely,

A. "An answer exists to a question X" means "There exists an alternative such that the risk associated with its selection is zero."

We must now make certain preliminary remarks on the meaning of the problem "How does the mind answer questions?" Here the problem of classification is most difficult, for history is not at all clear as to the modes of opinion on this subject. Some philosophies of history treat the matter in accordance with psychological concepts: we learn by sensation, or by reason, or by intuition, etc. Some emphasize the ontological aspect of knowledge, and state that we learn by correspondence of image and real, or coherence of concepts, etc. These classifications appear to be historically fruitful, but they do not appear to satisfy the conditions of logical exhaustiveness: Are sensation, intuition, and reason the *only* modes of learning? If so, what makes them exhaustive?

Rather than attempt to elucidate the concepts of epistemology which seem to be tied up with cultural meanings difficult to unravel and make consistent, we shall rather direct our attention to the experimental aspect of the problem. The experimenter faced with

the task of designing an experiment and the apparatus to carry it through is constantly on guard concerning the assumptions he makes in the design, in the use of the apparatus, and in the inferences he makes from the results. He is worried, we shall say, about the *presuppositions* involved in answering the specific question with which he is concerned. Is it always necessary to presuppose some information before one can set about answering a problem? If so, what kind of presupposition must be made? Or, in the language already adopted, is it necessary (1) to presuppose a knowledge of law in answering questions of fact, or (2) to presuppose a knowledge of fact in answering questions of law? We take these two problems as the basis of opinions on "how" the mind comes to know.

It should be noted that, just as we could not rely on common sense to make entirely unambiguous the concept of "fact," so we cannot rely on it to make unambiguous the concept of "presupposition." But the formulation of statistical methodology in Chapter II does suggest a manner of defining "presupposition" that will have the degree of precision required by the subsequent analysis. Accordingly, as a second basic principle of our inquiry, we shall assume that

B. "The answering of a question X presupposes the answering of a question Y" means the same thing as "Some one of the possible alternative responses to Y must be asserted in all the alternative responses to X."

For example, we have already suggested that "randomness" is commonly made a presupposition within statistical theory, and by this have meant that the assertion of randomness appears in every alternative hypothesis which we regard as a possible answer.

As a philosophical note on these matters, there is no claim made here that the term "presupposition" corresponds to the concept of the a priori, which is the subject matter of so many "metaphysical" studies, though we do feel that a far more profitable interpretation of Kant's thought can be made by the translation of *his* meaning of the a priori into the suggested language.

There are two aspects of the meaning of presupposition that are important for the classification about to be proposed. These aspects arise in the critical cases where a given question is supposed to be

unanswerable in the sense given above. If the answer to X does not exist, then shall we say that its (vacuous) answering demands presuppositions? Again, if the answering of X presupposes the answering of Y, but Y is actually unanswerable, then what about X?

To the first of these questions, it appears that we can make either sort of reply; the meanings of the terms involved do not seem to force us logically into either an affirmation or a negation. But the history of the problem has indicated the manner in which the reply should be made. If the question never receives a final answer, and if nevertheless the human mind keeps making attempts to find answers, then to characterize this struggle it has seemed reasonable to assert that presuppositions are constantly being made (though with futility). Accordingly, we add to the assumptions of our inquiry the assertion:

C. If the answer to X does not exist, then the answering of X (vacuously) presupposes the answering of all questions.

With regard to the second question, whether an unanswerable presupposition Y is sufficient to guarantee the unanswerability of a question X, the meaning we have assigned to presupposition seems to provide the kind of reply we must make:

D. If no response to Y is ever given without some risk, and if responses to X depend upon *some* answer to Y, then no response to X is ever given without risk.

Supposing now that "fact," "law," "answer," and "presupposes" are precisely enough defined, we can proceed to the classification as follows. Everyone who expresses an opinion on the fundamental problem as to what we know and how we know it must agree with all the presuppositions A, B, C, and D, and agree or disagree with each of the following propositions; and all the possible ways of affirmation and negation of these propositions represent a logico-historical classification of opinions in the sense defined above:

1. *The answering of any question of law presupposes the answering of at least some questions of fact.*

2. *There exist answers to at least some questions of law.*

3. *The answering of any question of fact presupposes the answering of at least some questions of law.*

4. *There exist answers to at least some questions of fact.*

The student of combinations will realize that the possible ways of affirming or denying four propositions are sixteen so that our classification of opinions appears tedious indeed. However, the presuppositions A–D, which we are accepting for the purposes of this classification, preclude certain possibilities on logical grounds. For example, no one can *consistently* assert the following set of three opinions:

1. All questions of law presuppose the answering of at least some questions of fact.

2. There exist answers to at least some questions of law.

Not-4. No questions of fact are answerable.

For by D, if there exist no answers to any question of fact, then any question whose answering presupposes the answer to a factual question must itself be unanswerable. In logical terms, assertion 1 and 2 *imply* assertion 4, so that 1 and 2 and not-4 is an inconsistent conjunction. Hence, the combination of any other assertion with these three will be an inconsistency; e.g., 1 and 2 and 3 and not-4, or 1 and 2 and not-3 and not-4.

Similarly, assertions 3 and 4 imply assertion 2, so that to assert not-2 and 3 and 4 is to assert a contradiction.

Again, no one can consistently assert:

Not-1. Some questions of law do not presuppose the answering of any questions of fact, and

Not-2. No answers to questions of law exist,

for by principle C, if 2 is denied then all law-questions vacuously presuppose the answering of any question. Hence no one can assert not-1 and not-2 together, and hence no one can assert not-1 and not-2 in conjunction with any other assertions. Similarly, not-3 and not-4 represent a contradiction.

On the basis of these logical results, we can construct a scheme of all possibilities. We use the logical symbolism of ab to stand for "both a and b are asserted," a' to stand for "not-a," and 0 to stand for an inconsistency. Where a given combination is consistent, a name has been given to correspond with the historical example. The defense of such correspondence will be the subject of the concluding paragraphs.

The sixteen possibilities are, then,

1 2 3 4	= Experimentalism
1 2 3 4′	= 0 by D
1 2 3′ 4	= Naive empiricism
1 2′ 3 4	= 0 by D
1′ 2 3 4	= Spinozistic rationalism
1 2 3′ 4′	= 0 by C and D
1 2′ 3 4′	= Relativism
1′ 2 3 4′	= Hegelian rationalism
1 2′ 3′ 4	= Statistical (Humian) empiricism
1′ 2 3′ 4	= Criticism (Kant)
1′ 2′ 3 4	= 0 by C and D
1 2′ 3′ 4′	= 0 by C
1′ 2 3′ 4′	= 0 by C
1′ 2′ 3′ 4	= 0 by C
1′ 2′ 3′ 4′	= 0 by C

I–II. The school we shall call "rationalism," and which we believe to have been the fundamental thesis of Spinoza and, in some of his writings, Leibnitz, is characterized by denying 1, and asserting 2 and 3. That is, the rationalist feels that laws are a priori to facts, that laws can be known. His attitude toward the answering of questions of fact is not clear-cut and gives two types of rationalism, depending on whether 4 is answered affirmatively or negatively. In seventeenth century rationalism, questions of fact are usually considered to be answerable; in Descartes and Spinoza, for example, the particular can always be found, and the classic example of the attempt to derive a fact from laws is the ontological proof for the existence of God. Within the Hegelian philosophy, however, the pure individual is excluded as an object of rational knowledge. The Cartesian rationalist based the knowledge of law upon intuition and the weakness of his position lay in the fact that intuition thereby became a sufficient condition for answering some questions. If a critical examination of intuition is permitted, then the principles of the critique must be granted, and must supposedly be granted intuitively, so that intuition still remains sufficient. The problem of replacing the immediacy of intuition by a methodology is suggested in one of his writings by Leibnitz, whose dream was that logic

alone would be a sufficient condition for all truth, though logic itself is presumably known intuitively. The rationalist, in effect, laid down the conditions of formal science (using geometry as a model); the failure of his method is the same failure that faces anyone who attempts to make abstract or formal science self-sufficient: We may become more and more precise concerning the rules of deduction and the criteria of consistency, but the ultimate criteria of precision appear to lie outside the realm of the formal science itself. It seems that if one would avoid an infinitude of "metalanguages" one must have recourse to nonformal method. In brief, what this story retains from rationalism is that formal science and intuition are necessary conditions for answering questions; what it discards is that formal science and intuition are sufficient conditions for answering questions.

III–IV. The antithesis of rationalism was empiricism: the empirical attitude is to assert 1, deny 3, and assert 4. That is, facts are fundamental, and a priori generalizations (laws) do not exist; the attitude here toward the answering of questions of law parallels the rationalist's attitude on questions of fact. Some empiricists (e.g., John Locke and in some instances J. S. Mill) seem to think that complete induction from facts is possible. This position, for obvious reasons, we prefer to call *naive* empiricism, and its weakness lies in supposing that a finite set of facts can be sufficient to establish a generalization. The more sophisticated empiricism attempts to reduce all questions of law to questions of fact; instead of asking, for example, whether all planets behave in accordance with Newton's Law of Universal Gravitation, let us ask about the probability[1] that this Law holds in the light of our present evidence, in much the sense that one asserts in biology the probability of a dominant characteristic in the second generation of hybrids. We can call both questions legal if we choose, and designate the answer as a law (the latter is called Mendel's Law), but it is actually a single value (the probability) that we are demanding, and according to the definitions of law and fact, the question is really factual. This opinion we call *statistical* empiricism, because the procedure of determining the required probability is statistical. What history had to abandon,

[1] Modern statistics sometimes uses the term "likelihood" in this connection.

with Kant, was the hope of empiricism to escape the necessity of a priori law, for as Kant showed the very meaning of factual experience demands a priori generalization; what history retained was the necessity of answers to questions of fact in the answering of any question, whether general or particular, and, in the case of statistical empiricism, the necessity of statistical methodology in the consideration of questions of law.

V. That the Kantian position was a synthesis of rationalism and empiricism is well recognized. Criticism, as expounded in certain passages of the first *Critique*, denies 1 and 3, and asserts 2 and 4; that is, *both* law and fact must be known before experience begins; there must be the immediate intuition of experience (fact) plus the "given" forms of experience (law) before a complete experience exists at all. What history has absorbed of Kant is his insistence on the interdependence of law and fact, and the necessity of intuition in all science; what it has abandoned is his position that factual and legal intuition are sufficient to answer questions in some cases: the modern mind no longer holds to the immediacy of sense data, nor to the particular forms of thought (Euclidean geometry and Newtonian mechanics) that Kant supposed were fixed for all time. It is strange that though Kant's contribution far outweighs his error, the present tendency is too often to disregard the former because we have come to recognize the latter.

VI. The antithesis of criticism is relativism, whose first beginnings are to be found in Maimon, but whose most complete expression lies in pragmatism (e.g., in Schiller especially, and to some extent in Dewey's *The Quest for Certainty*). The relativist position asserts 1 and 3, but denies 2 and 4; that is, relativism recognizes the necessity of a priori law, and yet finds that to *know* the a priori as true one must first know certain facts. Thus, to answer a question of fact, one must answer first a question of law, but to answer the latter, certain facts must be known. The relativist accepts the conclusion apparently implied in this situation, and argues that all truth is relative; truths are starting points for action, just as intuition is, and the particular "truth" an individual will take depends upon a number of factors, both sociological and psychological. Answers, in the absolute sense given above, do not exist. It would be

difficult to judge what history will take or reject in this position, since its process is contemporary. It will certainly take the relativist's emphasis on purpose. It also appears that there can be no denying the relativist's position regarding the truth of 1 and 3. Even a cursory review of experimental method convinces one that the taking of a single observation presupposes an entire theory, and yet that no theory can be said to have claim to validity until it has been put to the test of observations. Thus we cannot even *begin* to gather evidence (facts) to support or refute a theory without having a theory to start with, and yet what absolute right have we to start with a theory that has not yet been substantiated? If history does take this much from relativism, must it take the rest? Is scepticism a consequence of this much?

VII. The only remaining position is that which asserts 1 and 3, and yet retains 2 and 4; this position seems to be untenable, if not on logical grounds, then certainly on the grounds of common sense. How can we make fact depend on law, and law on fact, and yet finally give answers to both? It is difficult to propose a reply, because we have not made explicit enough the meaning of the phrase "an answer exists." We have purposefully not done so because the meaning was left vague in the history of philosophy; the meaning has usually been assumed to be known on intuitive grounds, and the demands for elucidation have not been great. But at the present point of philosophical progress these demands are made urgent: unless we are to admit the claims of a relativism, and grant a fundamental scepticism in all inquiry, then we must attempt to define what a nonscepticism would be like, and, since scepticism is based on the assertion that there are no absolute answers, the meaning of scepticism and nonscepticism evidently depends on the meaning of an existent answer.

I think our review of modern philosophy has shown that we cannot define the existence of (absolute) answers to questions as propositions that (a) can be found in a finite number of steps and (b) will satisfy the demands of the questions asked under all possible conditions; we cannot, I say, so define the existence of absolute answers without admitting the relativist's position that no absolute answers exist. For if modern philosophy has shown that propositions 1 and

3 must be asserted, then every question presupposes the answering of something else; this something else presupposes a third, and so on. The circle cannot be completed without ending up with the absurdity that a question presupposes its own answer.

It is suggested that the meaning of "existent answer" be identified with the meaning of the objective of all scientific endeavor, since the basic purpose of all science is to give answers to meaningful questions. Then he who asserts that answers can be given in a finite number of steps asserts that the objective of science is attainable, and hence is a "goal." In one sense, the "progress" of modern philosophy has been to show that the objectives of science are not attainable. But there is a viewpoint, typically modern, that to say an objective is unattainable is not the same thing as saying the objective does not exist. This viewpoint would insist that an objective that may be approached within any given distance, however small, exists, even though it can never be reached. If answers are not attainable in a finite number of steps, they may exist nonetheless, in the sense that they are endpoints (limits) of a progress that may eventually approach them within any given degree. Those ends that are unattainable but approachable in this sense Singer calls "ideals." The task that remains, the task of experimentalism, is to show how the answers to questions may be considered as ideals, in the light of propositions 1 and 3. That is, experimentalism, having broadened the meaning of "existent," has as its chief problem the definition of the concept of answer, under the assumption that all of the four propositions set down above are true. This we take to be the fundamental problem of the philosophy of science: to define "answer" in such a manner that answers exist even though propositions 1 and 3 are valid. It is a problem the groundwork for the solution of which has been given by Singer, who generalizes on the methods of measurement theory. The true measure of a given distance will be the limit of an infinite set of observations, all in "statistical control." When lack of control results, the scientist changes his theory, so that theory depends on observation, and yet no observation can be made without *some* presupposed theory. That the ideal exists, is a necessary postulate of all inquiry, and the set of all ideals is the "real world." Statisti-

cians call the ideal a "stochastic limit"; Singer, borrowing from Kant, calls it a *Grenzbegriff;* whatever designation is used, the ideal is the "true value," and all truth is ideal. Thus experimentalism absorbs the claim of rationalism for the necessity of formal theory, the claim of empiricism for the necessity of observation, the claim of criticism for the necessity of the interdependence of law and fact, and the implication of relativism for the necessity of redefining the object of scientific inquiry; out of all these necessities supplied by the history of thought, it constructs what it takes to be the *sufficient* condition for correct methodology.

By way of summary, it is of interest to consider the progress of modern philosophy here outlined from a different direction; since each position is a product of its age, and the product of the individuals of that age, one may consider each as a manifestation of either a sociological or a psychological "type."

Among the psychological types it is easy to recognize the following two: (1) there is the type which exhibits "lag" with respect to the choices of means for the solution of problems; that is, the type tends to select one pattern of behavior as absolute, and is unapt to change the pattern even though the purposes or conditions so change that the pattern is no longer the most efficient; (2) the type which exhibits "anti-lag" is apt to change even when a change of behavior pattern is not efficient.

It is the contention here that the history of modern philosophy represents essentially a contrast of these two types. At the beginning of modern experimental science, there was a clear-cut demand for self-evident "starting-points" which the early scientist did not feel obligated to investigate. This demand was strong, for a scientist faced with the necessity of investigating all his presuppositions would be inclined not to start; indeed, he would be at a loss as to what a starting point would be like. This demand for "beginning-points" in science was supplied by the rationalist and empiricist. As science developed, it became clear that a failure to investigate the so-called self-evident propositions of reason or sense often led to unreliable and meaningless results. The demand was met in part by criticism, which attempted to reduce the uncontrolled first beginnings to a minimum. But a tendency towards *lag* set in

within philosophy, and the well-entrenched rationalist, and especially empiricist, schools continued to argue the necessity of their types of first beginnings, long after this kind of philosophy had ceased to be efficient for the scientist. When the tendency towards lag is carried far enough, the means which were originally beneficial actually are detrimental. This, as we shall hope to show, is the present characteristic of rational and empirical philosophies of science; these philosophies, far from describing the conditions under which some questions of science are answered, actually fail to account for the adequate answering of *any* question the scientist raises.

The counter-movement to the lag exhibited by the rationalist and empiricist has been the anti-lag characteristic of relativism. The "liberalism" of the relativist takes the stand that since there are no first-beginnings for science, the correctness of procedure oscillates from individual to individual, or from social group to social group. No absolute criteria of "right and wrong" can be found either in science or morals for the relativist. This means that the choice of means to our ends changes as widely as possible; there is no point of reference which can stabilize this constant change in plans of action which characterizes the relativist. If relativism is carried far enough, it leads to an anarchism which denies any criteria whatsoever in the choice of means, or the solving of problems.

Modern experimentalism attempts to solve the antinomy of lag and anti-lag, by demanding on the one hand the existence of some "ideal" in terms of which scientific progress can be measured, while remaining completely methodical, and free from first-beginnings in the realm of present activities of science. The ends of science are in some sense "fixed" (though subject to investigation), but the means of approaching these ends are not.

We turn now to a detailed account of the dialectical process thus briefly summarized.

REFERENCES

1. Cowan, T. A., "Legal Pragmatism and Beyond," in *Interpretations of Modern Legal Philosophies* (Essays in Honor of Roscoe Pound), edited by P. Sayre, Oxford Univ. Press (1947).

2. Singer, E. A., Jr., *Dialectic of the Schools* (mimeographed), to appear in *Experience and Reflection.*

3. Singer, E. A., Jr., "Logico-historical Study of Mechanism," *Studies in the History of Science*, Univ. of Pennsylvania Bicentennial Conference (1941).

Chapter V Rationalism

Of all the philosophical faiths the reflective mind has been prone to feel toward its problem of truth-finding, the strongest seems to be its faith in man's powers of reasoning. Even the empiricist's faith in our sensory powers never succeeded in annihilating the primacy of reason in its own domain.

Our attempt here is to review the story of the rationalist's faith, and to examine critically its fundamental claim: that the process of reasoning is in some instances sufficient to establish certainty in science. The critical task we have set ourselves will demand a defining of "the process of reasoning," an explanation of its claim to certainty, and an examination of this claim. The purpose is not so much to portray a history of the growth of reason, as to use that history in clarifying what we must mean by the concept, and in illustrating the difficulties involved in the claim for the self-sufficiency of reason.

One way of regarding the early thought of Greece is to consider it as a dialectic, out of which grew the formulation of the fundamental laws of classical logic. To Parmenides it appeared that logical law must exclude all other law from nature; to Heraclitus it appeared that natural law must exclude all logic. To Aristotle, the synthesis was made by recognizing two aspects of nature: the formal and the material. He to whom the formal is all-important will recognize the primacy of the laws of logic set down in the *Prior Analytics*. But let him inquire into the behavior of natural bodies (matter as well as form), i.e., into problems concerned with change; he will discover that logic alone can never suffice to give answers. None can deny that if all *a* is *b*, and all *b* is *c*, then all *a* is *c*. This is a formal property

of all things. But this law is not sufficient to establish a single proposition about objects of nature; it can never of its own accord tell us that all men are mortal, that Socrates is a man, or that Socrates is mortal.

It is true that later history has sometimes been prone to call the process whereby we learn the truths of nature a process of logic or reasoning, and, to distinguish it from the formal process, has called it "inductive" logic. But whatever may be the justification for this terminology, the sort of logic it signifies has certainly not enjoyed the kind of faith we have ascribed to reasoning; men have rarely if ever exhibited an absolute confidence in their powers of analyzing the changing phases of nature. No, it is in the other kind of logic, the formal, that the feeling of certainty has been entertained, and it is this sort of reasoning that is to be the subject of our discourse. We shall divorce ourselves, then, from those problems that worry the experimental scientist, and attempt to characterize the type of truth-finding indulged in by the mathematician and formal logician. Whether we shall succeed in maintaining the separation we now so confidently make must depend on the sequel of our story.

The formal logic of the *Prior Analytics* was to remain practically unchanged for two millennia. The next important step in the ancient period consisted in showing how the methodology of logic could be extended to other domains. This was finally accomplished, through the aid of his mathematical forebears, by Euclid in the *Elements*. There was born the essence of the formal method in science, to act as the criterion of all subsequent exact mathematics. It is true that it was not the formal logic of Aristotle that was used as a basis for constructing Euclid's deductive system, but rather a sort of "visual-intuitive" reasoning. But Euclid's motivation was the same as Aristotle's. The particular application of the Euclidean method was geometry, but it was evident enough that the method could be applied to all fields of inquiry; when so extended, the result is a "formalization" of the given subject matter.

The Euclidean construct was made up of a set of axioms and postulates, by means of which certain theorems were deduced. The system had much the same property that Aristotle's syllogism

enjoyed; *if* one grants the postulates and axioms (premises), then the theorems (conclusions) must follow.

The relationship between Euclidean formal science and non-formal science was certainly not clearly recognized for many centuries; it took a history of mathematical confusion to bring us to our present attitude towards the purpose of formal science. The confusion in Euclid's mind is best seen in his definitions. In defining "point," "line," etc., Euclid apparently thought it necessary to explain these concepts in terms of observation, so that his reader could identify them with objects in the natural world. Such definitions never serve any function in his deductions; for example, nowhere does he make use of the fact that a point has no dimensions, or that a straight line "lies evenly with the points on itself."

The confusion in the post-Euclidean mind concerning the purpose of formal science is best seen in the history of Euclid's "Fifth Postulate." The Fifth Postulate was an assumption Euclid felt obliged to make in order to obtain the converse of one of this theorems. On the basis of what looked to be very simple geometrical assumptions, he had been able to show that if a transversal cuts two lines so as to make the alternate interior angles equal, the lines must be parallel (i.e., must never meet). But suppose the alternate interior angles are not equal; then must the lines intersect? However obvious it may have appeared that all such lines should meet, Euclid found it impossible to *deduce* this consequence within the system already developed. He therefore postulated this property.

This apparently facile method of getting around the difficulty worried his successors: the postulate was not simple in its statement, and did not enjoy the "obviousness" of the rest of Euclid's assumptions. In this objection to the Fifth Postulate, these early mathematicians were implicitly voicing a philosophy of formal science which was to flower into the rationalism of the seventeenth century; the thesis of this type of rationalism is that a perfectly constructed formal science is sufficient to answer all the meaningful questions the mind can pose. "Perfection" in this respect implied not only that the formal conditions of consistency and rigor of deduction were satisfied, but also that the postulates themselves were intuitively evident first principles. To a rationalist such as

Spinoza, the future of science must have begun to simulate a job of bookkeeping: simply to derive from established first principles all the truths of science. This outlook on science is characteristic of rationalist philosophies: that when the fundamental problems are answered, and the process of science completely analyzed, the remainder of our days can be devoted to following out the explicit instructions of science.

A survey of certain aspects of the traditional (seventeenth century) rationalism will serve to show how attempts were made to formalize science along these rigorous lines.

The rationalist's central problem is to describe how the scientist comes to know the first principles of all knowledge. One answer offered by historical rationalism to this problem is found in the works of René Descartes. Descartes set out to discover the primary principles of science by simply doubting all statements offered by science, where the criterion of doubt rested upon an ability to bring forth sufficient reasons why a certain proposition *might not* be true. In other words, the criteria of a primary principle will be its *indubitableness*. There is no need of our following the philosopher through the tortuous paths by which he came to cast doubt on practically every statement that mankind had habitually accepted, including such mathematical dicta as $2 + 2 = 4$. Suffice it to say that he finally discovered one statement the doubting of which was impossible, namely, the statement "I am thinking." To doubt this statement is impossible, since the very act of doubting is an act of thinking. Thus "I am thinking" must be true, and it is true because all acts of doubting are acts of thinking. Descartes' next step consists in the proof that he exists. We have shown, the argument runs, that "I think" is true; hence "I exist" must be true. Now a logical analysis of this argument shows that it contains implicitly another premise, for from the single premise "I think" alone we cannot infer "I exist"; the inference contains three terms ("I," "existent things," "thinking things"), and the logician does not recognize a three-termed Immediate Inference as valid. The only premise allowing us to conclude "I exist" from "I think" is the statement (in the form of a law, be it noted) "All thinking things exist." As to why we need accept these laws,

Descartes' answer seems to be that statements are indubitable if their declarations are clear and distinct and self-evident:

> After this I considered generally what in a proposition is requisite in order to be true and certain; for since I had just discovered one which I knew to be such, I thought that I ought also to know in what this certainty consisted. And having remarked that there was nothing at all in the statement "I think, therefore I am" which assures me of having thereby made a true assertion, excepting that I see very clearly that to think it is necessary to be, I came to the conclusion that I might assume, as a general rule, that the things which we conceive very clearly and distinctly are all true (6, pp. 101–102).

In effect, then, our knowledge of those laws that form the basis of all science is arrived at immediately, without any further presupposition; in other words, Descartes advocates the employment of an immediate intuition in the discovery of the answers to at least some questions. Further examples of arguments for the use of intuition in the apprehension of basic principles are the following:

A. The method of intuitionism is inherent in Plato's doctrine of the soul, though clear examples of his advocating this method are hard to find in the Dialogues. For Plato, the soul at birth contains all knowledge, though it is not aware of this fact. Experience simply acts as a reminder of principles once known in the life of the soul prior to birth. Plato was led to this theory (apparently) through the teachings of his predecessor Socrates; the Socratic method of truth or way of arriving at knowledge (the Socratic dialectic) consisted of a search for the true definitions of things by question-and-answer. Definitions of such a concept as "courage," say, would be proposed by Socrates' pupils; the master would then criticize the definition offered by presenting an example of a thing that was courageous but did not fall under the given definition, or a thing that evidently was not courageous but did fall under the definition. Evidently, Socrates presupposed that his audience was in agreement fundamentally on all these concepts, and for Plato at least, this presupposition implied that all the abstract ideas existed in the soul in the same manner. Since variable experience could hardly give rise to the uniformity of ideas required, the theory of innate ideas was a necessary consequence. The doctrine of innate

ideas seems to lead naturally to intuitionism, for the recollection of an idea once had or a principle once known would presumably be immediate, or at least unconscious, just as the recollection of a forgotten name seems to take place without conscious reflection on our part. Further, the one recollecting has no doubt of the truth of the principle he has recalled, a characteristic of the intuitive method. The doctrine of innate ideas, be it noted, is not a necessary condition of intuitionism.

B. Perhaps the clearest outline of the intuitive method and its place in the process of learning occurs in Spinoza's works. In his essay *On the Improvement of the Understanding* (14) Spinoza states that there are four possible methods (reduced to three in the *Ethics*) by which the reflective mind may answer the questions it raises.

(1) The method of "hearsay," sometimes called dogmatism: "By hearsay I know the day of my birth, my parentage, and other matters about which I have never felt any doubt" (14, p. 7). Dogmatism is essentially the truth-methodology which determines the answers to questions of fact and law by means of an appeal to an (external) authority; in the traditional logic the doctrine is classified under *ignoratio elenchi* (ignorance of the point of issue) as the Fallacy of *ipse dixit;* in the older schools, where the master's word was final, arguments were settled by an appeal to this word. The doctrine of dogmatism might fit in any of the schools we have outlined, so that we shall not consider it separately. As a matter of fact, there are very few cases where the method is actually employed in its "pure" form. For example, the modern theological doctrine of revelation is not necessarily a form of dogmatism, for appeals to the Bible to settle theological disputes are not made uncritically; it has first to be established (by a nondogmatic method) that the Bible *is* the revealed word, so that *basically* the doctrine of revelation rests on some other method than dogmatism. Similarly, the inclination to believe what we read in the papers is a product of the "cultural compulsive" that makes us accept the news columns as reliable. Spinoza finds the dogmatic method inadequate, for it "must always be uncertain, and, moreover, can give us no insight into the essence of a thing . . ." (14, p. 9). Spinoza's criticism is based on one of his criteria for an adequate

truth-method, namely, that it should provide us with an exact knowledge of our nature, and as much as is needful of nature in general (14, p. 5).

(2) The method of "mere experience" is the second candidate Spinoza considers for a truth-methodology; here knowledge is arrived at through "experience not yet classified by the intellect, and only so called because the given event has happened to take place, and we have no contradictory fact to set against it, so that it therefore remains unassailed in our mind" (14, p. 7). This method, which we shall examine in some detail later under the head of empiricism, Spinoza considers inadequate in that "its results are very uncertain and indefinite, for we shall never discover anything in natural phenomena by its means, except accidental properties, which are never clearly understood, unless the essence of the things in question be known first" (14, p. 10). Thus, mere experience might teach us that this table is brown, that white, this four-legged, that six-legged, but as to what a table is *essentially* (i.e., what the necessary properties are for a thing to be a table), mere experience can never tell us; indeed, the statement "This table is white" already presupposes the knowledge of what a table is, and hence mere experience presupposes for its simplest judgments some other method.

(3) The method of inference arises "when the essence of one thing is inferred from another, but not adequately; this comes when from some effect we gather its cause, or when it is inferred from some general proposition that some property is always present" (14, p. 7). Evidently, though this method "gives us the idea of the thing sought, and enables us to draw conclusions without risk of error, it is not by itself sufficient to put us in possession of the perfection we aim at" (14, p. 10). Since any inference presupposes something from which the inference proceeds, the presupposed proposition cannot itself be inferred in every case.

(4) The only adequate method, for Spinoza, is employed "when a thing is perceived through its essence; when, from the fact of knowing something I know what it is to know that thing . . ." (14, p. 7). This method, described in the *Ethics* as "intuition," is not the consequence of something else, but is immediate. To seek a method

that is, mediate, i.e., dependent on something else, is to seek the impossible: "In order to discover the best method for finding out the truth, there is no need of another method to discover such a method; nor of a third method for discovering the second, and so on to infinity. By such proceedings, we should never arrive at the knowledge of the truth, or, indeed, at any knowledge at all" (14, p. 10). In the *Ethics*, the intuitive method is described as follows: "He who has a true idea knows at the same time that he has a true idea, nor can he doubt the truth of the thing" (13, Prop. XLIII, Pt. II). Thus, the intuitive method grasps the truth with such force that no grounds for uncertainty remain. It has two properties, (a) it apprehends the truth of a proposition "immediately," without recourse to other knowledge, and (b) the apprehension is such that there remains no doubt of the truth in the thinker's mind, nay, not even a doubt that he knows the truth (nor a doubt that he knows he knows the truth, etc.).

Spinoza compares the proposed four methods by an illustration, designed to show the indisputable superiority of the last.

Three numbers are given; it is required to find a fourth, which shall be to the third as the second is to the first [e.g., 2, 4, 3 might be the given numbers, and we are required to find the number x such that $2/4 = 3/x$]. Tradesmen [using the method of hearsay] will at once tell us that they know what is required to find the fourth number, for they have not yet forgotten the rule which was given them *arbitrarily, without proof*, by their masters; others [using the method of experience] construct a universal axiom from their experience with simple numbers, where the fourth number is self-evident, as in the case of 2, 4, 3, 6; here it is evident that if the second number be multiplied by the third, and the product divided by the first, the quotient is 6; when they see that by this process the number is produced which they knew beforehand to be proportional, they infer the process always holds good for finding the fourth number proportional. Mathematicians, however, [using the method of inference] know by the proof of the nineteenth proposition of the seventh book of Euclid what numbers are proportionals, namely, from the nature and property of proportion it follows that the product of the first and fourth will be equal to the product of the second and third: still they do not see the adequate proportionality of the given numbers, but, if they do see it, they do see it not by virtue of Euclid's proposition, but intuitively, *without going through any process* (14, pp. 8–9).

Spinoza's insistence that the intuitive method is an immediate apprehension of the *essence* of a thing, be it noted, places him in the school of rationalism, for the knowledge of essence is always the knowledge of law ("*all* men are rational," "*all* triangles have angles summing to two right angles," etc.). Thus, since knowledge of essence is obtained without presupposing anything (in particular, without presupposing knowledge of fact), Spinoza asserts the postulate that at least some knowledge of law presupposes no knowledge of fact, and in his rejection of the method of experience he seems clearly to assert that all knowledge of fact presupposes a knowledge of law, and thus makes the two fundamental assertions of rationalism.

C. The distinction that is so important for most rationalists between the *contingency* of the truths of sensation and the *necessity* of the truths (laws) of intuition is brought out clearly in the writings of William Whewell: "There is another mode in which the distinction of the two elements of knowledge appears; namely, in the distinction of *necessary* and *contingent* or *experiential*, truths. For these two classes of truths, the difference arises from this; the former (necessary truths) are truths which, we see, must be true: the latter are true, but so far as we can see, might be otherwise. The former are true necessarily and universally: the latter are learnt from experience and limited by experience" (16, Bk. I, p. 58). The following is Whewell's statement of the rationalist's position that the knowledge of laws (i.e., necessary and *universal* truths) does not presuppose the knowledge of facts: "Now with regard to the former kind of truths (the necessary truths), I wish to show that the universality and necessity which distinguish them can by no means be derived from experience; that these characters do in reality flow from the ideas which these truths involve; and that when the necessity of the truth is exhibited in the way of logical demonstration, it is found to depend upon certain fundamental principles (Definitions and Axioms) which may thus be considered as expressing, in some measure, the essential characters of our ideas" (16, p. 58). These fundamental truths are nontautological: "It may be said that the assertion that three and two make five merely expresses what we mean by our words; that it is a matter

of definition; that the proposition is an identical one. But this is by no means so. The definition of five is not three and two, but four and one. How does it appear that three and two is the same number as four and one? It is evident that it is so; but *why* is it evident? — not because the proposition is identical; for if that were the reason, all numerical propositions must be evident for the same reason. If it be a matter of definition that three and two make five, it must be a matter of definition that 39 and 27 make 66. But who will say that the definition of 66 is 39 and 27?" (16, p. 59). No, such necessary truths cannot be doubted because "we cannot conceive them to be otherwise," a common criterion of the self-evident for the intuitionist. Truths thus intuitively known are *laws:* "Necessary truths must be *universal* truths. If any property belong to a right-angled triangle *necessarily*, it must belong to *all* right-angled triangles" (16, p. 64).

It now becomes our task to examine the intuitive method critically. We may do so by appealing to either of two basic criteria of a general methodology, namely, by showing that the intuitive method leads to contradictions, or that it leads to no answers at all. To substantiate the first claim, that the intuitive method leads to contradictions, we have only to point out the all too numerous cases, found in the history of science, of men who have taken certain principles to be self-evident and indubitable, principles that later investigations have indicated are false. The flatness of the earth's surface must have seemed obvious to some, as did the proposition that the earth is stationary while the sun and stars move around it. So popular did the rationalist doctrine become that in the eighteenth century the deductive method was applied even to such sciences as economics and jurisprudence. It seemed "obvious" enough to the laissez-faire economists that a society which is allowed to develop of its own accord will obey certain principles, and for them economics became deductive in character, its sole task being to determine the consequences of fundamental and self-evident principles, principles that nowadays seem to be anything but confirmed or evident in actual societies. In the same way, jurisprudence was made to begin with "self-evident" principles concerning the law (*ius naturale*), the particular laws of a given land (*ius gentium*)

being derived therefrom; but the most superficial study of jurisprudence is enough to convince one that the nature of *ius naturale* has undergone considerable change in jurisprudential writings, and hence that the "obvious" fundamentals of law of yesterday are not only no longer obvious but appear wrong today.

Perhaps the best example of an historical refutation of the complete adequacy of the intuitive method is to be found in the history of geometry. The attempts to establish criteria of "well-founded" first principles inevitably led to failure: either the criteria themselves had to be proven on the basis of other self-evident principles, or an infinite regress or circularity of basic principles resulted. Actually, it was the mathematical investigations into the Euclidean Fifth Postulate which led to the fundamental difficulties involved in an absolute rationalism, for these investigations ultimately produced a non-Euclidean formal system.

The history of the Fifth Postulate was a series of attempts to "prove" it within the original Euclidean system. Such attempts usually ended in showing the postulate to be equivalent to some other proposition which the mathematician took to be "obvious" enough not to demand further proof. The postulate is equivalent to asserting that the sum of the angles of a triangle are equal to two right angles, that three angles alone do not determine a triangle, that asymtotic straight lines do not exist, etc. The fact that history nowhere stopped at any of these reductions shows a continuous, though for the most part unconscious, recognition on the part of geometers that "obviousness" cannot be a criterion of correctness in formal science.

In the early eighteenth century, in the work of Bolyai and Lobatschevski, a non-Euclidean geometry was constructed in which the Fifth Postulate did not appear and could not be deduced. This geometry was formally correct, in that the postulates were consistent and the deductions were carried through correctly. This discovery marked the dawn of a new era in the philosophical attitude towards formal or deductive science; it was no longer possible to assert that formal science was sufficient to establish the truths of nature, since, for a given set of concepts, many contrary formal systems could be constructed. It is true that by a proper translation of the concepts

of one formalization, a consistency between the various deductive sciences of geometry could be shown. Thus, the non-Euclidean plane geometry that denies the Fifth Postulate can be shown to hold in a solid Euclidean geometry, by defining a non-Euclidean line as a geodesic on a certain surface; a similar transformation shows that Euclidean plane geometry holds on a horosphere in a solid non-Euclidean space. But these transformations cannot be made if the concepts of geometry are supposed to have specific meanings within the world of nature. If a "straight line" is taken to refer to a particular aspect of the natural world (say, the path of light *in vacuo*), then the formal systems of geometry become contrary descriptions of nature, and we are led to inquire which is the "correct" one. Such an inquiry could hardly be conducted on formal grounds alone without involving us in an infinite regress or finite circularity.

The intuitionist might argue, in the face of these many historical illustrations of the breakdown of his method, that in these cases the faculty of intuition was used "uncritically." But the necessity of such critical use of the faculty implies that anyone convinced of the self-evidence of a principle must always test his conviction "critically." We then inquire whether the critical test is intuitive or not. If intuitive, then it must be tested again by another critical test, and this test by another, and so on to infinity. If not intuitive, then, some method is more fundamental and our task becomes one of searching for this more general method.

Again, we may argue that the intuitive method is actually inapplicable. We are asked, when presented with the answer to a certain question of law, to decide whether the answer is "self-evident." If the intuitive method is meaningful, it must explain under what conditions an answer is self-evident; but this it cannot do, for when a proposition is intuitively known, it must also be immediately known that the proposition is self-evident; by "immediately known" is meant that no presuppositions, no further method, are required to establish the self-evidence. There is no *method* employed to show that a proposition is intuitively known; he who knows it, knows that he knows it. For one not so gifted as to know things in this way, or one having a tendency to doubt all things, the intuitive method

must then be meaningless, since for him there is no way of applying its fundamental demand. Not "feeling" the self-evidence of any proposition, he has no way of answering questions by the intuitive method. The method, then, becomes restricted to those who have intuitions, and is not *generally* applicable.

All this may be summarized by raising a question which from the point of view of the historical survey we have had to ignore: what is the meaning of "intuition"? A witticism of Ambrose Bierce's, which is more truly to the point than most wit, defines the "self-evident" as "evident only to one's self." The comment is turned to fruitful use if we regard intuition as a personality characteristic: the types of intuition we display measure in some manner our own individual characteristics, and the characteristics of the culture and social groups to which we belong. Thus viewed, intuition is not a general basis for a complete methodology, and actually the use of the intuitive faculty can be studied within an experimental psychology or sociology. We shall want to return to this viewpoint when a more adequate definition of experimental method has been found. It should be emphasized very strongly that, in abandoning the use of intuition as a *sufficient* condition for truth-finding, we have not denied its function as a *necessary* condition. If we should demand of an adequate method that it employ no intuitions in any form, then science could never make any progress in its search for truth.

The inadequacy of intuition as a sufficient condition for truth discovery forced the rationalist position to the narrower viewpoint that at least *formal* science can be developed and completed without reference to nonformal (experiential) criteria. The newer (contemporary) rationalism thus claims the autonomy of abstract science: provided "sufficient" care is exercised in the selection of postulates and in the deduction of theorems, a formal theory may be completed, though its completion does not infer that any description of natural events has been given.

To take a simple example of the revised rationalism (borrowed from Leibnitz), suppose we are asked to determine the truth of the proposition

$$2 + 2 = 4.$$

Instead of using the intuitionist's appeal to self-evidence, let us analyze the terms involved, "2" and "4." The number "2" may be defined as follows:

$$2 = 1 + 1.$$

To show that $2 + 2 = 4$, we also define "3":

$$3 = 2 + 1.$$

Hence: $4 = 3 + 1 = 2 + 1 + 1 = 2 + 2$, by means of our definitions. The beauty of this proof is that it follows from the *definitions* of the concepts involved; in other words we have (apparently) shown that the truth of $2 + 2 = 4$ *follows from the definitions of the numbers.* We have, then, only to inquire about the truth of these definitions; but such an inquiry is needless, since definitions are simply the arbitrary conventions by which we designate certain ideas by certain symbols or words. Thus, instead of writing "$1 + 1$," we write "2," instead of saying "thing used for the purpose of writing in ink" we say simply "pen." Is it possible to "*refute*" a definition? One might object to the awkwardness of certain symbols or words, but he would not be denying the definition. The ideal of formal science then becomes clear: we are to make every true statement become deducible from the definitions of the concepts involved. Postulates and axioms disappear from the perfect deductive science; all that remains are definitions and the logical laws of deduction. Leibnitz actually tried to carry out this method for the science of jurisprudence by simply setting down the definitions of the terms of jurisprudence and deducing therefrom certain basic principles.

Because of the weaknesses of the symbolic method in Leibnitz's time, his ideal of a perfect, truth-by-definition deductive science could not be carried out. Indeed, in the simple illustration given above, there is a serious error. For to derive the truth of the equation

$$2 + 1 + 1 = 3 + 1$$

from the equations $2 = 1 + 1$ and $3 = 2 + 1$ we must assume (at least) the validity of the law

$$(2 + 1) + 1 = 2 + (1 + 1),$$

for we must reassociate the elements of the addition in order to write $2 + 2$ for $(2 + 1) + 1$.

Such difficulties have been overcome by modern investigations into deductive theory and the foundations of mathematics. The "rigor" the ages had found in Euclid's treatment of geometry became nothing but a loose form of reasoning under the scrutiny of nineteenth-century mathematics; the deduction of theorems to be found in the *Elements* often makes implicit use of propositions whose "truth" is nowhere formally asserted. Further, as we have mentioned, many of Euclid's basic concepts were defined in an empirical fashion, and the definitions nowhere form a part of the theory. The newer logic, or rather the newer theory of formal science, revised the nature of the formal construct briefly as follows: A set of undefined notions are presupposed, which are taken as abstract concepts; i.e., they are abstracted from any content other than that ascribed to them by the postulates. In terms of these undefined concepts, certain other derived concepts may be defined. Postulates in the form of propositions or propositional functions are asserted. A set of "Rules of Procedure" is given describing the manner in which deductions are to be made. Theorems are deduced on the basis of the definitions, postulates, and rules only.

Such is the revised account of the manner in which formal theory is to be developed, the resulting systems apparently make no reference to the world of experience, and the concepts and meaningfulness of the systems do not depend on experiential intuitions. If one inquires into the value of such abstract theoretical systems, the answers vary; there is the point of view that the systems are valuable in themselves, in much the sense that a great painting is; on the other hand, there is the point of view that they are valuable only when applied, their proper application being a matter of scientific intuition. The question of value, however, is not so much our concern here as is the question whether the rationalist's claim of autonomy for abstract science is justified.

As the theory of abstract science developed, it became clear that there were really two aspects to be considered: (a) one might investigate a problem entirely "within" the realm of the theory itself (for example, one might attempt to derive a certain theorem of the system), or (b) one might investigate a problem "about" the entire theory (Is the theory consistent? Can certain propositions

be deduced within the theory? And so on). The latter type of problem is essential in the completion of a formal theory, since the problem of completeness itself evidently belongs to this type. It can be shown, however, that the language we use to discuss a theory cannot be a "part" of the theory itself. For this reason a metalanguage is essential, a language which is adequate to discuss the general problems relating to a given theory. For example, the criteria of deduction must borrow from the metalanguage: the Rules of Procedure, which purport to tell us exactly how theorems are to be deduced within the system, usually involve metalinguistic concepts; the Rule of Substitution specifies the manner in which certain symbols of the system are to be substituted for others in the deduction of theorems.[1]

With the realization of the necessity of a language which "talks about" a certain formal system, there has resulted in the journals and elsewhere discussions that remind one of the scholastic disputes of the thirteenth and fourteenth centuries, even in those days in disrepute. For example, it seems to be important in these discussions to distinguish, within the system, between the "use" and the "mention" of a symbol, between many meanings of "meaning," between different connotations of the term "truth," etc. It might shock the unprepared reader to find in Chwistek and Hetper's "New Foundation of Formal Mathematics" (3) a statement such as this, "the sign C is the fundamental element of our expressions. It has no meaning at all." Evidently, the authors have a restricted meaning of "meaning" in mind, namely, that the symbol does not stand for some known usage, and that its use in the system does not imply implicit principles.

The first question that comes to mind in connection with metalinguistic considerations is whether these can be formalized themselves. In the sense in which we have used the term "metalanguage" so far, the answer is usually in the affirmative, and the formalization commonly goes under the name of "syntax." Within this more or less exact framework, the metalinguistic account of a formal system

[1] In many contemporary treatments of formal science, the older form of the Rules has been replaced by certain devices which tend to simplify metalinguistic considerations (4).

can be given, and the vagueness which confused the procedure of deduction in the classical systems can be removed. Instead of discussing such problems as those of "completeness" and "consistency" in an intuitive manner, we can presumably translate these problems into precisely formulated problems within the formal syntax, in the same manner as classical arithmetic or geometry was able to translate certain vaguely stated problems into a language capable of giving exact solutions. This property of formal syntax has proved very fruitful in our understanding of certain formal systems; for example, by the use of formal syntax, Gödel (7) was able to show that all the "classical" formulations of arithmetic are incomplete in the sense that there exists an arithmetical proposition X such that the (metalinguistic) statements "X is provable" and "X is false is provable" are both false within the formal syntax. The newer scholasticism has at least surpassed the older in the precision with which it treats its terms.

The rationalist viewpoint with respect to the newer developments in abstract theory is that formal method is self-completable. It asserts that the exact methods of handling problems within formal theory, or within the syntax of a formal theory, can be extended to the handling of all problems arising out of formal method.

The failure of the rationalist viewpoint can be demonstrated in at least two ways. The first is founded on formal grounds alone. Formal incompleteness must exist, in the sense that not *all* problems within syntax are solvable (12, pp. 310 f.). The equanimity with which many formalists face this indeterminism of formal method reminds one of the easy nonchalance with which many physicists face the so-called indeterminacy of physical science. For further discussion, see Chapter XIII.

The second argument against the rationalist thesis is a more general one. It consists in pointing out that at no stage in the construction of formal theory has intuition been avoided as a sufficient condition for answering certain questions. Such a thesis undoubtedly requires elucidation of the meaning of intuition, and in view of this requirement cannot be stated with precision until a better formulated science of psychology shall have explained the term. For the present, we can only translate the term "intuitive" into terms that

will require a better defining, but whose common meaning may help to elucidate. We say that a step in a process is intuitive if an exact method is not given whereby the step is taken; equivalently, a step is intuitive if one is not completely "conscious" of the nature of the step (assuming, without further discussion, that "consciousness" and "method" are equivalent in denotation).

None who have worked within formal science can deny the empirically obvious fact that all steps in a deduction are partially "intuitive"; how the next step is to be taken, and once taken whether it is legitimate are problems that are never decided in a completely methodical fashion. Those who have not devoted their energies to mathematical fields may find it difficult to accept this "obvious fact." For them, the realm of mathematical thinking may have appeared as a sought-for haven of errorless investigation, and the so-called exact procedures of mathematics may have seemed to be the one example we have of true precision. But the fact that this belief is based on an ignorance of the field, and that the belief is shared by few who are acquainted with the foundations of mathematics, suggests that the term "precision" has been left vague in the minds of nonmathematicians. It seems evident enough that if mathematics is to be a precise science, then not only must the chance of making a wrong deduction be zero, but one must also be able to *show* that this chance is zero. At present it can be shown by many examples that the chance is not zero, and also at present there is no clear-cut, nonintuitive method known for demonstrating that a given deduction must be correct.

The refutation of modern rationalism can perhaps best be summarized by considering the problem of the distinction between analytic and synthetic propositions. Those who claim that logic is "empty," or that logic contains only tautologies, or that logic does not "talk about" the world, have necessarily proposed for themselves the Kantian problem of defining unambiguously the term "analytical judgment." Suppose first we try Kant's own definition, that analytical judgments are those in which the meaning of the predicate is "contained in" the meaning of the subject (9, Intro., IV). There is, of course, the Hegelian paradox concerning this definition: that one cannot assert that the predicate's meaning is

completely contained in the meaning of the subject, else it would not be predicate any longer. But a more comprehensive, and comprehensible, objection to this meaning of the analytic is that it does not specify accurately enough the meaning of "contained in."

Within modern logic the meaning of this term has received more precise defining through the methodology of formal science. We would now like to say that all judgments are analytic which "follow from" the principles of logic, according to certain well-defined rules of procedure. To make such defining accurate, we must state what we believe the principles of logic to be, we must formalize these principles by symbolic techniques, and we must state precisely the rules of deduction that hold within our system. The results of such a construction will be a system of analytical judgments.

Now this much defining of analytical judgments appears trivial until we have shown what "logic" is, what constitutes adequate as opposed to inadequate symbolism, and finally what constitutes correct formal deduction. We would like to say that logic is that which abstracts completely from the content of empirical science; we would like to say that the adequacy of symbolism is defined by rules whose meaning is independent of any empirical connotation; we would like to say that the correctness in matters of deduction may be made explicit without recourse to empirical judgments. If we could say all these things, we would have shown that logic is "empty," that analytical judgments are possible, and perhaps that within formal science we can attain a degree of certainty that the empirical sciences can never provide.

The refutation of modern rationalism thus consists in showing that none of these desires, for an empty logic, a nonempirical symbolism, a purely formal deduction, can be fulfilled. To show this requires a further specification of terms, however, for even the modern formulation of the divorce of the analytic and the empirical is vague in its meaning. Do we mean that nothing the empirical sciences contribute can in any way influence our opinions on the meaning of logic, or symbolic techniques, or formal deduction? Certainly this would be a naive translation of the claim to emptiness in logic, in view of the enormous influence the sciences have had upon the work of symbolic logicians. Do we mean that *it is possible*

eventually to perfect a system of formal logic without recourse to the empirical sciences? This seems to be the only translation that is adequate, and yet such a translation flies in the face of much evidence. It would mean that in some sense we can decide the principles of logic independent of the psychological and sociological conditions that normally influence our behavior, i.e., we would have to assert that in deciding whether a given judgment "abstracts from the content of all experience," we do not have to take into account the personality bias inherent in all other scientific investigations. For if such bias did play a role in our decision, then it would not be true that the science of logic had avoided the necessity of empirical methodology. Again, if we took seriously the proposed divorce of the formal and empirical sciences, then we would mean that the process of combining symbols in a certain fashion is again independent of personal and social conditions, i.e., that the rules of procedure can be stated so as to avoid errors due to personal bias in their application. And yet all the evidence collected by the linguists in their study of language formation shows clearly the tendency of individuals to combine sounds and symbols in a certain way, and almost invariably to avoid certain combinations.

If formal logic cannot escape the error due to a personal or social bias, then we are forced to deny that we have *attained* a purely empty logic, or an absolutely exact deductive system. We have not denied that a general logic may be taken as an ideal to be approximated indefinitely. The concept of an empty logic is as meaningful certainly as the concept of an exact distance; yet both are unattainable without presupposing a great deal about the rest of nature.

Thus, even though we admit the possibility of an exactly constructed formal logic, we must also admit that its attainment is impossible unless we presuppose some information about the individuals (and their society) who study the problem. And this admission is tantamount to denying the rationalist thesis in its modern form: the answer to any question of law (even the tautologies of logic) does presuppose the answering of at least some questions of fact.

The refutation of rationalism implies that the future progress of

formal techniques must be in nonformal or experimental domains; that it is the task of the future psychologist (or sociologist) to analyze the type of mind that patternizes, and to explain the properties of the intuition that is now a necessary element of formal theory. It may be, as Quine asserts (12, p. 291), that at present the proof of a theorem is discoverable in general only by luck; but if the future is to consist in removing the chance-variables associated with formal science, such "luck" should be investigated by an experimental analysis of the psychology of the mathematical mind. The "good luck" of which Quine speaks is not a random good luck, to be found in anyone who happens to attempt a proof; it is a "good luck" we find in a particular psychological type, and in more or less degree within the type. What are the characteristics of such minds? What is it about such personalities that enables them to follow a certain pattern? Exactly what sort of problem will have to be answered in order to eliminate the chance-element of formal science will evidently depend upon the future of an experimental psychology of personality.

The discussion of this chapter has been general in the sense that we have considered the role of presupposition with respect to the answering of questions of law within any science. We now return by way of illustration to the special problem of inference to see in what way a rationalism would attempt to answer certain of the basic problems we raised in Chapter III.

A typically rationalist solution of the meaning of randomness occurred in the early applications of probability theory. A set of chips are said to be drawn "at random" from a bowl if each chip has the same chance of being drawn. Now how shall we determine this equiprobability? The rationalist solution was a principle of "insufficient reason": the chances are equiprobable if we can supply no reason to the contrary. This negative sort of application of a rational intuition is well known to lead to certain paradoxes, e.g., Bertrand's Paradox (11, p. 114); the paradoxes are actually a result of a lack of method in the application of the principle: how are we to decide a lack of sufficient reason? And yet, if the principle of insufficient reason proves inadequate to define randomness of observation, then what criteria shall we use? Can we assert that the

chips are distributed at random if the bowl is mixed in a certain "haphazard" fashion? Or should we not rather try to formalize the meaning of randomness so that we have a precise mathematical definition of its meaning? This latter attempt corresponds to the developments of modern rationalism we have previously discussed. The attempt consists in regarding the problem of randomness to be a purely formal one within the structure of mathematics; the task is to define a random sequence as a special type of number sequence. Such is the purpose of von Mises' Kollectiv (11). The aim here is to exhibit the formal character of randomness; the meaning of the term is apparently divorced of any empirical content, and we should thus be able to formulate its meaning unambiguously within formal theory.

Our refutation of this application is based, not so much on the question how such a formal definition could be put to practical use, as on the question whether a purely formal definition could ever be given. And the answer to the latter problem is the same as the refutation we have already given of modern rationalism: the demand for "purity" in formal matters cannot be met without recourse to empirical evidence. We can regard a purely formal definition of randomness to be a mathematical ideal, the attainment of which presupposes an empirical methodology. If such an empirical method is presupposed, and if the empirical method itself demands a concept of randomness, then we will require a *nonformal* definition of randomness before we can attain the end of a purely formal definition.

A rationalist attitude towards the basic probability theory can also be found in the literature, e.g., in (8). The theory is regarded as a purely formal construct of a language of probability. The refutation of this viewpoint scarcely needs repeating: a purely formal probability theory will have to presuppose a theory that is not purely formal (i.e., is nonformal), granted that empirical science requires for its methodology an empirical definition of probability.

The thesis we have developed in this chapter, which negates the rationalist claim for completeness in any domain of science, plunges us into another problem whose answering is not simple. If formal theory presupposes some kind of empirical method for its comple-

tion, then have we not involved ourselves in a vicious circularity? For, as the next pages will show, empirical method must in general rely upon formal science for its completion. Nor will the remark that the distinction between formal and nonformal science is only a relative one suffice to remove the paradox. And yet the main thesis of this essay is that we must be content to live with such circularity, and the main purpose of the essay is to suggest how one may so live without conceding the demands of scepticism that such circularity commonly implies. Before proceeding to the outline of a "circular theory" of inquiry, we next examine other viewpoints that attempt to escape circularity by means of an appeal to the immediate data of sensation.

REFERENCES

1. Bonola, R., *Non-Euclidean Geometry*, translated by H. S. Carslaw, Open Court (1912).
2. Carnap, R., *Introduction to Semantics*, Harvard Univ. Press (1942).
3. Chwistek, L., and Hetper, W., "New Foundations of Formal Mathematics," *Journal of Symbolic Logic* (1938).
4. Curry, H. B., "Revision of the Fundamental Rules of Combinatory Logic," *Journal of Symbolic Logic* (1941).
5. Descartes, R., "Meditations on First Philosophy," in *The Philosophical Works of Descartes*, translated by E. S. Haldane and G. R. T. Ross, Cambridge Univ. Press (1911).
6. Descartes, R., "Discourse on the Method of Rightly Conducting the Reason," in *The Philosophical Works of Descartes*, translated by E. S. Haldane and G. R. T. Ross, Cambridge Univ. Press (1911).
7. Gödel, K., "Uber formal unentscheidbare Sätze der PM und verwandter Systeme," *Monatshefte Mathematik u. Physik* (1931).
8. Halmos, P., "The Foundations of Probability," *American Mathematical Monthly* (1944).
9. Kant, I., *Critique of Pure Reason*, Meikeljohn translation, Bell (1872).
10. Leibnitz, G. W., *Philosophical Works*, translated by G. M. Duncan, New Haven (1908).
11. Mises, R. von, *Probability, Statistics and Truth*, Macmillan (1939).
12. Quine, W. V., *Mathematical Logic*, Norton (1940).
13. Spinoza, B., "Ethics," in *Spinoza: Selections*, Ed. by J. Wild, Scribners (1930).

14. Spinoza, B., "On the Improvement of the Understanding," in *Spinoza: Selections*, Ed. by J. Wild, Scribners (1930).
15. Tarski, A., "Der Wahrheitsbegriff in den formalistieren Sprachen," *Studia Philosophica* (1936).
16. Whewell, W., *History of Scientific Ideas*, Appleton (1858).
17. Wittengenstein, L., *Tractatus Logico-philosophicus*, London, Paul (1922).

Chapter VI Naive Empiricism

The feeling of dissatisfaction the reader may have felt with the rationalist's answers to our problems is no new one; the history of epistemology is replete with illustrations of such objections to the "rationalist" way of things and searches for something much more "practical." The antinomy seems to be a comparison between the thinker who, immovable, contemplates the mysteries of the universe from his armchair, and the practical fellow who goes out into the world and learns about things through hard knocks. It is this latter type of reflective mind we now wish to study.

He who would forsake his armchair for the world of hard knocks in order to learn the ways of things is usually called one who "learns by experience"; or, perhaps better, by observation. He is the thinker who uses his senses to decide on questions the mind may raise. He refuses to believe that "there are in the understanding certain innate principles, some primary notions, . . . characters, as it were, stamped upon the mind of men, which the soul receives in its very first being, and brings into the world with it" (2, p. 37). No, for him the mind at birth is like a white paper, void of all characters, without any ideas. How do we gain all the ideas of whiteness, hardness, sweetness, thinking, etc., we have? To this the empiricist answers, in one word, from experience, "In all that our knowledge is founded, and from that it ultimately derives itself" (2, p. 122). And what is this fountainhead of our knowing, this experience? It is really a composite of two things:

First, our senses, conversant about particular sensible objects, do convey into the mind several distinct perceptions of things, according to those various ways wherein those objects do affect them: and thus

we come by those ideas we have, of Yellow, White, Heat, Cold, Soft, Hard, Bitter, Sweet, and all those which we call sensible qualities; which when I say the senses convey into the mind, I mean, they from external objects convey into the mind what produces there those perceptions. This great source of most of the ideas we have, depending upon our senses, and derived by them to the understanding, I call sensation.

Secondly, the other fountain, from which experience furnisheth the understanding with ideas, is the perception of the operations of our own mind within us, as it is employed about the ideas it has got; which operations, when the soul comes to reflect on and consider, do furnish the understanding with another set of ideas, which could not be had from things without; and such are Perception, Thinking, Doubting, Believing, Reasoning, Knowing, Willing, and all the different actings of our own minds; which we being conscious of and observing in ourselves, do from these receive into our understanding as distinct ideas, as we do from bodies affecting our senses. This source of ideas every man has wholly in himself; and though it be not sense, as having nothing to do with external objects, yet it is very like it, and might properly enough be called internal sense. But as I call the other sensation, so I call this Reflection, the ideas it affords being such only as the mind gets by reflecting on its own operations within itself. . . . These two, I say, viz. external material things, as the objects of sensation; and the operations of our own minds within, as the objects of reflection; are to me the only originals from whence all our ideas take their beginnings (2, pp. 122–124).

The account of how the mind, starting with nothing, comes to obtain its many and complex ideas and beliefs is a long and tedious one for the empiricist, but essentially it is a description of a construction; the *material* of the construction is comprised of the facts of experience, the arrangement is due to certain functioning powers of the mind (abstraction, memory, reason, etc.; these powers, incidentally, lead us to deny the original analogy between a white paper and the mind). The concept of the material of all knowing is what characterizes the empiricist, for it consists of certain "immediate sensations (or reflections)." A few examples will suffice to make clear the nature of such experiences. If I were to show you a piece of paper and ask you whether it *looked* red to you, I would be raising a question whose answering, for most empiricists, is immediate; you would know immediately and without further analysis whether you were seeing red or not. It is important to notice that the ques-

tion whose answering is immediate is concerned with one's own impressions, and not with the nature of the object observed; we do not ask "*Is* this a red piece of paper?" Regardless of what the object may be like in itself or to others, your answer will be right if you simply respond by stating what you observe. Similar examples of immediately known facts are familiar enough: there can be no doubt of your feeling pain when the dentist touches a sensitive place with his drill; having the pain and the knowledge that you have it are identical for the empiricist. Note again that it is no refutation of the empiricist's position to point out that people are frequently "mistaken" about their pains; the mistake arises only in their attributing the wrong cause to what they feel, not in their being mistaken about the feeling itself. Thus a man who has recently had his leg amputated may feel pain where pain can no longer exist, but an error would occur in his thinking only if he inferred the existence of a leg from the circumstance of having the pain.

A brilliant flash of light, a loud noise, a bitter taste, a strong smell, a rough-feeling surface, are all sensations the having of which leaves no doubts (for the empiricist) that one has had them. All of these examples can be put in the form of questions of fact since the subject is singular in each case: "Do you hear a loud noise?" "Do you observe a red patch?" "Do you have a painful sensation?" These questions of fact are answerable immediately, without presupposing anything else, for apparently one need draw on no previous knowledge to decide, for example, whether the sun appears bright or not. Hence all empiricists take for granted the characterizing postulate: "Some questions of fact can be answered without presupposing the answers to any questions of law." We note also that the answering of the questions of fact of the sort described is wholly a private one; no one but he who has the sensations can possibly decide (or approach the decision to) the correct answer concerning the nature of these sensations. Each man knows what passes in his own mind, and no man can ever have adequate knowledge of the private ideas of another. The world becomes divided into the "public" and the "private"; the public includes those things and movements all men share together — the stars and their movements, the earth and its structure, the animals and their

nature, etc.; the private contains those thoughts and ideas of a man's mind that are his own possession, his own way of looking at things. You and I might have "public knowledge" that the light of the sun takes about eight minutes to reach the earth, but just how that light looks to each one of us is a private matter, is a piece of knowledge not adequately transferable.

It needs no especially acute analyst to sense that the explanation of the immediacies of sensation and reflection as we have described them requires more clarification, and, as we shall see, certain empiricists have gone farther than we have done by reducing such so-called immediacies as the sensation of red to more ultimate ones. This further analysis we shall outline later; for the present it is sufficient to notice that according to the fundamental assumption of empiricism any analysis concerning the nature of the immediately knowable must stop somewhere with something no longer analyzable, with a something the having of which and the knowledge of having which are identical. The central problem of all empiricism becomes one of describing the manner in which knowledge (especially knowledge of law) is constructed out of these "blocks" of sensation and reflection.

For the naive empiricist, who holds that the knowledge of some laws is attainable after we have answered a sufficient number of questions of fact, the chief task of scientific methodology is to describe how the mind can arrive at these laws. The problem he sets himself is by its very nature a difficult one; he must show how we can answer with certainty questions of law that hold for more than the finite number of cases observed; he must show how these answers can be given when all we know are a finite number of facts. Philosophy has borrowed from logic a term to designate this central problem of naive empiricism. The logician defines the process whereby from certain assumptions we obtain additional statements (theorems), the process of "deduction"; he then designates as "induction" an inverse process, whereby, a certain group of statements being given, a set of assumptions is constructed which are sufficient to deduce the original group. Thus Euclid used an inductive process on the many geometrical "theorems" discovered by his mathematical predecessors; he constructed a set of postulates from

which all these theorems could be deduced; the Italian mathematician Peano, among others, performed a similar task for arithmetic. Indeed, most deductive sciences are constructed by an inductive process, the formalist having in mind at the beginning certain statements he wishes to be able to deduce. A special case of the "inductive process" is used by the empiricist; the statements from which the process starts are always facts, the data of sensation and reflection, and the statement with which it ends is usually a "law." The law thus induced "explains" all the observed facts, that is, if we grant the law, then the facts observed follow deductively as theorems. For example, suppose that we have observed (with Archimedes) that: (a) a certain lever balances if equal weights are at the ends and the fulcrum is placed in the center, (b) another lever balances if weights of 1 and 3 pounds are placed at the ends and the fulcrum is placed one-fourth of the way from the heavy end, etc.

Then we may induce a law sufficient to explain these observed facts: a lever will always balance if when weights P_1 and P_2 are hung at the ends, the fulcrum is placed so that the ratio $P_1 : P_2 = d_2 : d_1$ is maintained (d_1 being the distance of P_1 from the fulcrum, d_2 the distance of P_2).

Or, again, with Galileo we may observe certain facts about the behavior of a ball rolling down an inclined plane:

a) after the first second it travels 1 foot (approximately),

b) after the next second the distance travelled is 4 feet,

c) after the third second the distance travelled is 9 feet, etc., and of another ball on another inclined plane,

d) after the first second the distance is 2 feet,

e) after the next second the distance is 8 feet,

f) after the third second the distance is 16 feet, etc.

We come to induce the law that s, the distance travelled, is always equal to some constant times the square of the time ($s = kt^2$), the constant depending on the slope of the inclined plane.

The reader is familiar enough with similar though no doubt less mathematical inductions he must have made: the advent of cold weather, the symptoms of colds, etc.

The induction of a law is called "complete" if the law induced is certain to be true, i.e., if only one law is adequate for the explanation

of all the observed events (or if only one law will enable us to deduce all the facts). The naive empiricist postulates that such complete inductions are possible, and his central problem is to show what method must be employed to make such inductions.

Aristotle seems to have had to face the naive empiricist's difficulty, however he might differ from most examples of the school in other respects. For Aristotle, knowledge of universals (which we have designated knowledge of law), though possible, always follows upon knowledge of particulars, and he keenly felt the problem of explaining how the mind attains the former sort of knowledge. Put briefly, and hence to some extent inadequately, Aristotle seems to have felt that the mind had the ability to extract the essential from the accidental in its observations of things, and from these essentials it constructs its laws. Thus, from observations of many living things, we come to realize that, though color, location, size, etc., are all accidental to their being alive, nutrition and reproduction are essential, and hence we induce the law that all living things perform these two functions. A similar grasping of the essential is common in our everyday life: we realize that certain types of food are essential for living, certain are not; we realize that certain methods are essential for accomplishing our ends, certain are not. The clearest (though none too clear) explanation of Aristotle's concept of scientific method appears, perhaps, at the end of the *Posterior Analytics* (1, Bk. II, Chap. 19):

> When one of a number of logically indiscriminable particulars has made a stand, the earliest universal is present in the soul: for though the act of sense-perception is of the particular, its content is universal — is man, for example, not the man Callias. A fresh stand is made among these rudimentary universals, and the process does not cease until the indivisible concepts, the true universals, are established: e.g., such and such a species of animal is a step towards the genus animal, which by the same process is a step toward a further generalization.

The process whereby the mind gradually passes from the particular to the general (i.e., the "inductive" process) Aristotle calls "intuitive," and thereby absolves in a word the inherent difficulties of a complete empiricism. To the reflective mind, inquiring how one may pass from a finite collection of data to a generalization

covering all cases of a given type, it is not enough simply to describe the basis as intuitive, unless one is willing to defend an uncritical application of intuition to all investigations.

The modern empiricist feels obliged to replace the intuitive method by one more in accord with the empirical tradition of relying on sensation for the growth of knowledge. If we are to construct laws out of a certain finite collection of facts, it is evident that the choice of facts is all-important. For example, if one were not careful as to his choice of numbers, he might come to the conclusion that all numbers were even, since he could find as many numbers as he could wish to substantiate his theory. Again, a bad choice of observed facts might lead one to suppose that toads caused warts, that black cats were unlucky, that the phases of the moon correspond to the changes in weather, and so on. We must be sure that the observed facts so vary that there can be but *one* adequate explanation of them all, i.e., one law sufficient to explain them. This means that the methodology we prescribe must indicate what observations are to be made before we induce the law of nature explaining them. Such was the problem set by J. S. Mill in his *Logic* (3). For all practical purposes, the solution was supposed to have been given by his Canons of Induction.

The Canons of Induction are a set of methods for inducing the most important types of nature's laws, those of causality. The problem was one of setting down the method by which the experimenter may discover whether two events are causally connected. In effect, for the empiricist, two events, A and B, are causally connected if the appearance of A always concurs with the appearance of B. Thus, we say that a magnet causes a nail to move, since every time the nail and magnet are placed in a certain position the nail does in fact move. Causal statements, be it noted, are always in the form of laws, since they describe what *always* occurs.

Our task now is to make the correct observations to show that the event X is causally related to the event A. According to Mill there are five methods at our disposal (3, Bk. III, Chap. 8).

Canon I (Method of Agreement). Suppose the event X is observed to occur when the event A occurs and also when B, C, D, E, etc., are occurring (since no two events can occur completely iso-

lated). Now we are required to change all the "surrounding" events B, C, D, E, etc., but to leave A and to observe whether X still occurs. If X still occurs when everything but A has been changed, then X and A are causally related (A is the cause or effect of X).

One simple example will suffice to make the method clear. If you suspect that eggs are bad for you, i.e., if you wish to determine whether certain feelings of discomfiture are causally related to eating eggs, then the Method of Agreement prescribes that you observe whether the painful feeling follows the eating of eggs, bacon, toast, coffee, and also follows the eating of eggs, potatoes, milk, and orange juice. Other examples of this method will be given later, but for the present we should note that Canon I alone is hardly sufficient to establish a unique law explaining the observed facts. If X occurs when A, B, C, D, and E occur, and X also occurs when A, F, G, H, and I occur, *one* explanation may be that A and X are causally connected, since from this law we can deduce the observed facts, but a *second* explanation would be that B and X are causally related and also F and X are causally related (but that A and X are not). In the above example, the cause of our illness may be the bacon in the first case, the milk in the second. The second explanation is also sufficient to allow us to deduce the observed cases, and it is clear enough that to multiply the number of examples under the Method of Agreement will never give us a complete induction unless we know all the variables, for if X occurs also when A, J, K, I, L occur, and when A, M, N, O, P occur, etc., then possibly J and X, and M and X are causally related.

The second method reads as follows:

Canon II (Method of Difference). Suppose, as before, that X occurs when A, B, C, D occur; now we change *only* A, to E, say, and see whether X occurs when E, B, C, D occur. If X *fails* to occur, then A and X are causally related. (A is the cause, or an indispensable part of the cause, or the effect of X.)

The Method of Difference prescribes that we are to try leaving eggs out of a meal that has previously caused us pain and observe whether the illness still follows. If *not*, then we can assert that the eating of eggs is causally related to our illness. Notice that the Method of Difference does not establish anything if it fails; if X

occurs when A, B, C, D occur and X also occurs when E, B, C, D occur, A and X may still be causally connected. This might mean that A was sufficient for X, but not necessary. Thus, the first canon is really a test of whether A is *sufficient* for X, while the second canon is a test of whether A is *necessary*.

The third canon attempts to supply the necessary and the sufficient conditions:

Canon III (Joint Method of Agreement and Difference). If two or more cases in which X occurs have only A in common, while two or more cases in which X does not occur have nothing in common except the absence of A, then A and X are causally related.

For example, "the instances in which much dew is deposited (on a surface) . . . agree in this, and, so far as we are able to observe, in this only, that they either radiate heat rapidly or conduct it slowly: qualities between which there is no other circumstance of agreement, than that by virtue of either, the body tends to lose heat from the surface more rapidly than it can be restored from within. The instances, on the contrary, in which no dew, or but a small quantity of it, is formed, and which are also extremely various, agree (as far as we can observe) in nothing except in *not* having this same property" (3, p. 484). Consequently, some causal relationship between this property of materials and the existence of dew may be inferred. It will be noted that the Joint Method imposes a much stronger condition than the Method of Difference: the *absence* of A has to be the *only* thing a set of circumstances has in common; presumably, this means that every other aspect of the closed system of nature within which we are working occurs at least once and not always. The application of the Joint Method would therefore seem to be very restricted.

Mill's Fourth Canon depends for its application on the remaining ones. If we have discovered certain causal relationships between some events, then there is the possibility of discovering other causal relationships by means of

Canon IV (Method of Residues). "Subduct from any phenomenon such part as is known by previous inductions to be the effect of certain antecedents, and the residue of the phenomenon is the effect of the remaining antecedents" (3, p. 460).

Good examples of this method occur in astronomical experiments, where the effects of the heavenly bodies on one another are more or less precisely determined. It is known, for example, that the planets cause certain "aberrations" in the orbits of the other planets, which, if "unmolested," would travel in the path of an ellipse; further, the effect that any given planet has on another's orbit can be calculated with great precision by Newton's Law of Universal Gravitation. Now we make observations on the planets' orbits and discover, after having taken into account all the aberrations caused by the known planets, that certain aberrations still remain. We "subduct from the phenomenon of the orbit such part as is known to be caused by certain antecedents," i.e., the other planets, and still find that there remains something unaccounted for. The Method of Residues tells us that these remaining effects must be due to certain antecedents, the exact mass and position of which are given by Newton's Law. Turning a telescope in the required direction (as did the discoverers of Uranus), we observe the planet which by this method must be the cause of the residual aberrations.

One other method remains. Its significance can be made apparent by an analogy: We imagine a number of pieces of fine wire laid on the floor but covered over, except for their ends. Then to find which ends are connected, we simply pull one of the ends and observe which one moves at the other end. Nature might well be conceived in a similar manner; we observe simply the end points of two processes causally connected, as the striking of the hammer and the sound of the bell. The scientist's task is to see which "ends" are connected, and, when possible, he should use

Canon V (Method of Concomitant Variations). If variations occur in K whenever variations occur in A, then K and A are causally related.

There are many simple examples of the Method of Variations. An electric light is inferred to be causally related to a switch if the light goes on and off as the switch is shifted back and forth; the seasons of the year are causally connected with the variations of the angle of the axis of the earth to its orbit, since seasonal variations vary as this angle varies. The study of the cause of business cycles is conducted by trying to correlate certain observed social

and economic variations with the variations of factors of business. This canon was one of the first formulations of modern correlation analysis.

By means of these five canons Mill hoped to establish a method whereby, from the direct observations of a finite number of facts, the experimenter could infer universal laws of nature. In point of fact, Mill himself did not think the required inference could be made without some possibility of error, so that there was never a complete induction in an absolute sense, but we have classed Mill as a naive empiricist (one assuming that questions of law are answerable in terms of answers to questions of fact) because the exact error involved does not seem to be an important problem for him (as it is for other empiricists); he even seems to feel (3, Bk. III, Chap. XXI), that the general law of causality in nature is sufficiently confirmed to make the risk of a wrong decision negligible. There are answers to questions of law that can be made with zero risk.

Many illustrations of Mill's methods are to be found in the literature of the philosophy of science, especially in the biological sciences. The famous Postulates attributed to the bacteriologist Robert Koch are restatements of Mill's Canons. These are designed to provide a method of determining whether a particular disease (observable by certain symptoms) is caused by a particular parasite. Koch's Postulates are usually stated somewhat as follows:

1. The organism should be found in all cases of the disease in question, and its distribution in the body should be in accordance with the lesions observed.

2. The organism should be cultivated outside the body of the host, in pure culture, for several generations.

3. The organism so isolated should reproduce the disease in other susceptible animals (4, p. 788).

The first and third postulates are forms of the Method of Agreement, that the occurrence of the disease X take place whenever the occurrence of the parasite A takes place, no matter how the environment may change. The second postulate is an attempt to obviate a difficulty of the Method of Agreement, namely, that other things in the diseased body might cause the disease by changing A, so that the original parasite A was not the cause of X.

It is true that Mill's Canons formed an important landmark in the history of experimental inference; in a sense, they were the first formulation of the principles of experimental design to form a basic part of the modern methods of inference discussed in Chapter II. Our criticism of the Canons is directed toward the claim Mill makes for them: they are supposed to provide us with a causal principle on the basis of certain empirical facts. Actually, it is not very difficult to show that the application of any one of the five Methods *presupposes* a specific causal law about the natural world. That is, had not the experimenter already assumed a certain type of causality in nature, he would never be able to apply the Methods, and hence never would "induce" causality from his data.

In the first Method we are requested to make "changes" in the accompanying events in order to ascertain whether X still occurs. Practically every term of the method requires clarification; e.g., what is meant by a change? Since not *every* concomitant circumstance can be changed (we do not seek to change the structure of the moon when investigating the cause of arsenic poisoning), we must know *what* events are to be changed.

Evidently, we change just those events we think are also likely to have caused the result. But then we have presupposed that certain things do *not* operate as causes, and hence have presupposed at least a partial account of nature's causal principles. Again, in order that the Method of Agreement have any value, we must change the concomitant events to those we think are unlikely to cause X by themselves, and this principle of change will demand presuppositions concerning causality. The Method of Difference fares no better. We cannot hope to change "only" A in a nature which is itself changing constantly; we can keep "constant" only certain aspects of the natural world, namely, those aspects we believe are important. Those we believe to be unimportant (such as weather, say, or sun spots) we do not attempt to control. But the solution of the problem of what should and what should not be controlled presupposes a knowledge of the causal principles that operate to influence the event X. Similar remarks apply to the remaining Methods. Indeed, these remarks apply to any of the modern techniques of experimental design. The "validity" of a certain

design depends upon just such presuppositions concerning what should and should not be controlled, i.e., upon the correctness of our knowledge of the causal principles operating in nature.

In the light of this criticism, it becomes necessary to restate the empirical position. Instead of a search for laws based upon a complete induction, an attempt is made to measure the degree of confidence we can have in a certain generalization. In that this degree or measure of confidence is factual, the empiricist will have succeeded in abstracting the specific (and therefore meaningful) from the general. We turn now to this attempt on the part of statistical empiricism.

REFERENCES

1. Aristotle, *Posterior Analytics*, Oxford, Clarendon Press (1926).
2. Locke, J., *Essay Concerning Human Understanding*, Oxford, Clarendon Press (1894).
3. Mill, J. S., *A System of Logic*, 8th Ed., Longmans, Green (1872).
4. Topley, W. W. C., and Wilson, G. S., *Principles of Bacteriology and Immunity*, Williams and Wilkins (1941).

Chapter VII Statistical Empiricism

The naive empiricist made it his central problem to show how the experimenter might induce certain laws of nature from a finite number of observed facts at his disposal. The induction was for all intents and purposes "complete"; that is, the law induced by the method could suffer no exceptions, and one predicting the future on the basis of such an induced law could be "reasonably" sure of success. The postulate that such a complete induction is possible is that assumption which characterizes the naive empiricist, and we have shown in what sense the belief in complete induction is "naive" and cannot form the basis for an adequate description of scientific methodology.

David Hume marks the transition in history from an attempt to "prove" laws from facts to the position that the only meaningful assertions we can make about laws are assertions about our degree of confidence. Hume, a member of the school of British empiricism, like his colleagues could see no other avenue to truth and knowledge than that provided by our senses. But unlike his predecessors, Hume asked himself a question as simple in its meaning as it was significant in its consequences: granted that the only evidence I can possibly have for my belief in any statement is past experience, what evidence do I have for believing, for example, that when I strike this bell with a hammer a sound will arise?

So obvious a question could hardly have escaped Hume's predecessors, but they failed to realize that the most obvious answer to it was also an inadequate answer. One would feel inclined to say that the bell *will* ring because it *has* rung on all previous occasions. Now we know that it *has* rung on previous occasions by the evidence of

our senses. What logic forces us to make the jump to the future? The only premise sufficient to allow us to make this awkward leap would be one to the effect that what has occurred repeatedly in the past without fail must occur also in the future. But the empiricist's only evidence for granting such a premise would be past experience; and even granting that in the past this premise has held, what evidence do we have that it will hold in the future? The apparent impossibility of providing such evidence on empirical grounds led Hume to deny that our predictions of the sort described could have any sound epistemological basis. Instead, questions concerning causal principles must be reduced to questions concerning an individual's typical reaction to a set of repeated circumstances. The essential relationship between a cause and its effect cannot be known, for no set of facts could ever be sufficient to provide the true essence. But we can observe by direct observation the *habitual* tendency of minds to make a jump from a certain observed event to its expected consequence. Hume's description of these "conditioned reflexes" of the human mind follows:

Similar objects are always conjoined with similar. Of this we have experience. Suitably to this experience, therefore, we may define a cause to be an object, followed by another, and where all the objects similar to the first are followed by objects similar to the second. Or in other words, where, if the first object had not been, the second never had existed. The appearance of a cause always conveys the mind, by a customary transition, to the idea of the effect. Of this also we have experience. We may, therefore, suitably to this experience, form another definition of cause, and call it, an object followed by another, and whose appearance always conveys the thought to that other. But though both these definitions be drawn from circumstances foreign to the cause, we cannot remedy this inconvenience, or attain any more perfect definition, which may point out that circumstance in the cause, which gives it a connexion, with its effect. We have no idea of this connexion, nor even any distinct notion what it is we desire to know, when we endeavor at a conception of it. We say, for instance, that the vibration of this string is the cause of this particular sound. But what do we mean by that affirmation? We either mean that this vibration is followed by this sound, and that all similar vibrations have been followed by similar sounds: or, that this vibration is followed by this sound, and that upon the appearance of one the mind anticipates the senses, and forms immediately an idea of the other. We may consider the relation of cause and

effect in either of these two lights; but beyond these, we have no idea of it (4, Sec. VII, Pt. II, pp. 631–632).

The suggestion implicit in Hume's philosophical analysis is that we first specify what we can know on the basis of the immediate data of sensation, and that we then attempt to reduce the traditional problems of philosophy and science to these ultimate knowables. In effect, to adopt empiricism is to adopt a certain criterion of meaningfulness, and to adopt empiricism is to reject all questions (such as questions regarding the true essences of things) which cannot be reduced to these ultimate sense data. In order to understand the revised empiricism of Hume, we have first to examine in more detail the empirical criteria of meaning.

The problem of meaning is really as general as the problem of truth. We have taken for granted that the definition of truth can be given by presenting a method whereby certain questions can be answered. Now a certain question is said to be "meaningful" if there is a method whereby I can answer it. Thus the question "Are there any boojums in Africa?" is meaningless until a definition or meaning is given the term "boojum," and this meaning is given by providing a method for determining whether any given object is a boojum or not.

Since the problem of meaning and the problem of truth are interdependent, we should expect to find that each solution of the truth-problem has its own solution of the meaning-problem. And this is the case. For the rationalist, the only meanings terms may have are those provided by the definitions and postulates of a formal system. For the empiricist, terms gain their meanings through the "hard facts" of experience. For example, he takes the meaning of such a question as "Are there any white men in Burma?" to be given if certain "empirical" tests are presented whereby we can determine whether an object is a white man or not, and whether it lies in Burma or not. These empirical tests are simply questions whose answers are known immediately (e.g., "Is this object white?" might be such a question). If a given question cannot be reduced to such tests, then it is meaningless; that is, science cannot answer it one way or the other.

The empiricist's philosophy of the meaningful is no new doctrine; indeed, its statement was a common one in Greek and medieval philosophy. In modern philosophy it received its strongest support at the hands of John Locke and the succeeding English empiricists, Berkeley and Hume, and in France, Auguste Comte.

In contemporary philosophy the empiricist's position has received a new impetus at the hands of the "operational" school and the logical positivists.

Operationalism seems to be no more than an application of the empiricist's definition of meaning to certain terms within experimental science. For example, consider the "notorious" terms of physics, "mass" and "force"; most of the classical physicists and philosophers had failed to give these terms any precise empirical meanings. Beginning with Mach in his *Mechanics*, philosophers of science have urged a more careful analysis of these and like terms common in the language of the physicist. For the operationalist, this analysis must proceed by reducing the terms to certain "operations," whose meaning is simple and derived from some immediately knowable experience:

> Whatever the nature of the concept, whether it involves the physical operations of the laboratory or the "paper and pencil" operations of one or another kind of construct, we must demand that the operations be such that they can be unambiguously and straightforwardly performed (1, p. 124).

True, ambiguity can never be removed entirely, but

> Observation of what we do in meeting the situations of experience shows that as a matter of fact there are certain operations about whose performance we have no hesitation. . . . In foundation studies in which one wants to secure the maximum awareness of what one is doing and the maximum security that he is not involving himself in contradiction, one would do well to use only concepts whose meaning is found in such unambiguously performable operations. For in this way the description of the situation . . . reduces to the description of an actual experience, namely, the performed operations, and actual experience is not self-contradictory . . . (1, p. 124).

The empirical doctrine of meaning has also formed the basis of modern positivistic doctrines. For Carnap, the most general pre-

scription of scientific empiricism as to scientific meaning is that a concept has meaning if there is a rule stating the conditions under which it is to be applied. The concepts of "reducibility" and "confirmability" were developed for the purpose of making specific the form of these rules (2). Reducibility for Carnap has a logical connotation: a certain proposition "reduces" to another if the confirmation of the latter entails confirmation of the former. Thus reduction in this sense has a formal character, and is designed to show a formal relationship between two propositions; for example, the confirmation that your heart or brain is in the room may reduce to the confirmation that your living body is there, since the latter (presumably) entails the former. But now the pertinent question is where the process of reduction ends. "Obviously," says Carnap, "the reductions must finally come to predicates for which we can come to a confirmation directly, i.e., without reference to other predicates" (2). In effect he proposes that statements of the form "The space-time point *a* is red," "*b* is hot," etc., be regarded as directly confirmable sentences.

On the basis of operationalist, or positivist, criteria of meaning, the empiricist must regard questions of law in the same manner as any other questions he may raise; that is, he must show how such questions can be reduced for their answering to questions of fact. The problem of law can be solved only by setting up a method which shall be reducible to the fact-finding method.

In the light of our criticism of naive empiricism there are but three ways of regarding questions of law.

1. We may deny that any amount of fact-finding can ever give us the right to make predictions regarding the future, so that even though a given event (e.g., the melting of ice on a stove whose surface is over 200° Fahrenheit) has occurred repeatedly in the past and never failed to do so, we cannot say whether it is more likely than not to occur again. We call this position "sceptical empiricism"; it takes the sole task of science to be one of describing accurately the past. For the sceptical empiricist, questions regarding the future are meaningless; there is no way of reducing the question, "Will the ocean have tides in 2980?" to questions answerable by an immediate experience in the present. Says Hume:

Let the course of things be allowed hitherto ever so regular; that alone, without some new argument or inference, proves not that, for the future, it will continue so. In vain do you pretend to have learned the nature of bodies from your past experience. Their secret nature, and consequently all their effects and influence, may change, without any change in their sensible qualities. This happens sometimes, and with regard to some objects: Why may it not happen always, and with regard to all objects? What logic, what process of argument secures you against this supposition? My practice, you say, refutes my doubts. But you mistake the purport of my question. As an agent, I am quite satisfied in the point; but as a philosopher, who has some share of curiosity, I will not say scepticism, I want to learn the foundation of this inference. No reading, no enquiry, has yet been able to remove my difficulty, or give me satisfaction in a matter of such importance (4, Sec. IV, Pt. II, pp. 606–607).

The sceptical empiricist, be it noted, is not a true sceptic, since he does assert that at least certain questions (the immediately known "hard facts") are answerable.

Once we have admitted that these questions concerning the past and present are answerable, and these only, the only problem of science becomes clear: that of describing in as simple and convenient terms as possible the observed events of the past. "Science tells us nothing regarding the future," says the sceptical empiricist, "it simply gives the best description possible of the past, where 'best' means 'simplest.'" The meaning of "simplicity" is required, of course, to make the position complete, and such definition must be given on empirical grounds. Karl Pearson suggests one such criterion in the following passage:

By the formation of conceptions, which may or may not have perpetual equivalents in the sphere of sense impression, the scientist is able to classify and compare phenomena. From their classification he passes to formulae or scientific laws describing their sequences and relationships. The wider the range of phenomena embraced, and the simpler the statement of the law, the more nearly we consider that he has reached a "fundamental law of nature." The progress of science lies in the continual discovery of more and more comprehensive formulae, by aid of which we can classify the relationships and sequences of more and more extensive groups of phenomena. The earlier formulae are not necessarily wrong. (They are what the mathematician would term "first approxi-

mations," true when we neglect certain small quantities. In Nature it often happens that we do not observe the existence of these small quantities until we have long had the "first approximation" as our standard of comparison. Then we need a widening, not a rejection of "natural law.") They are merely replaced by others which in briefer language describe more facts (8, Chap. III, Sec. 10, pp. 96–97).

Pearson seems to suggest as criteria of simplicity "brevity of language" and "comprehension." Whether these criteria can actually be reduced to purely empirical confirmation is certainly questionable. We shall have occasion to return to the problem of simplicity of law later, for it reappears in all the schools.

Sceptical empiricism, because of the brevity of its claims, can only be criticized from the point of view of a general criticism of all empiricism: whether even the simplest description can be given without presupposing a causal structure in the world. We shall return to this general criticism in later chapters.

2. The second alternative open to one denying that questions of law are answerable without error is to assert that one may decide whether a law is "more likely than not" to hold in the future. One is to make this decision on the basis of the evidence at hand. The present alternative declares that one may do so without making any method explicit, i.e., he may do so on "intuitive" grounds. Since the method becomes partially intuitive and displays no characteristics of intuitionism that we have not already discussed, we shall not further consider this alternative (the alternative of "intuitive" empiricism).

3. Last, there is a method whereby in the light of past evidence we may employ an explicit method of making certain "probability" answers to questions of law. This last position, which is that most commonly accepted by the empiricist, we shall call "statistical" empiricism, since its problems can be reduced to the traditional problems of statistics.

The problem of the statistical empiricist is to assign probability numbers, as measures of our confidence in a certain law. These probabilities are to be measurable in terms of direct observation. Now the "law" in which we are interested need not hold in every case; if it fails to hold in a certain proportion of cases, then our

expectancy will be thereby modified. A universal law is therefore only a special case of any sequence of repetitious events; in this special case our expectancy of a certain repetition is almost certainty, or, in the probability scale, is nearly 1. This suggests that the probability of a law's holding, be measured in terms of relative frequencies of observed events. This suggestion has been worked out in some detail in the so-called Frequency Theory of probability.

The basic concepts of the Frequency Theory are the following: (1) a "reference-class," R, comprising the "universe" of things with which the inquiry is concerned (e.g., the class of people thirty years old on January 1, 1910); (2) a subclass of elements, A, belonging to R (e.g., the class of people in the United States who were thirty years old in 1910 and who were still living in January, 1920).

Now let n represent the number of elements in R, and a the number of elements in A. Then we define the *relative frequency* of A in R

$$fr(A, R) = \frac{a}{n}.$$

For example, if the number of people aged thirty in the United States on January 1, 1910, was 250,000, and 210,000 are still living on January 1, 1920, then the relative frequency of such survivals is $\frac{21}{25}$.

Now suppose we enlarge our reference-class indefinitely (e.g., we take *all* people aged thirty in 1910, then all people aged thirty at any time, and let A = those who have survived the ten years). We then define the *probability* of any random occurrence of an A in R as

$$p_A = \lim_{n \to \infty} fr(A, R) = \lim_{n \to \infty} \frac{a}{n}$$

if such exists. (The probability of an occurrence of A in R is the limit of a/n as n "approaches infinity.") Obvious objections to this definition will occur to the reader. In the first place, no amount of experimenting can ever make the number of elements in the reference class "approach infinity," and it is a well-known arithmetical principle that the behavior of a sequence of values for a finite number of these values is independent of the limit of an infinite sequence of which these values are a part. For example, the finite sequence

$\frac{1}{5}, \frac{1}{6}, \frac{1}{7}, \frac{1}{8}, \frac{1}{9}, \frac{1}{10}$ seems to approach 0, but if we append this sequence to the infinite sequence $1, \frac{3}{4}, \frac{5}{8}, \frac{9}{16}, \frac{17}{32}, \cdots$, the limit is $\frac{1}{2}$. Hence, no matter how large we make n, the behavior of the resulting values of a/n gives us no indication of the *true* probability, for the number of these values is finite, and the true probability requires an infinite sequence.

To give an adequate answer to this objection, we must introduce a very important theorem of probability-theory discovered by an eighteenth century mathematician, Bernoulli. Its *inexact* statement may be given as follows: "If the true probability that a certain event, E, will occur is given by the number p, then with a larger and larger number of trials the frequency of E's occurrence will with greater and greater assurance become closer and closer to p." [1] Now the true probability is not known, but if Bernoulli's Theorem is assumed, then with an increase in the number of cases (i.e., enlargement of the reference class) we should become more and more certain of our guesses as to the true probability.

The theorem may be put more precisely: Let the true probability of E be p, let ϵ and η be two numbers, as small as you please, but less than 1 and greater than 0 ($0 < \epsilon, \eta < 1$). Then there exists a number n such that if we take n trials, the chance that the relative frequency of E differs from the true probability p by an amount less than ϵ is $1 - \eta$. We say the relative frequency approaches p "stochastically."

An example will help to make the theorem clear. Suppose our reference class comprises the results of the throws of a coin; also suppose that the coin and its throw are "perfect" so that the true probability that the coin lands heads is exactly $\frac{1}{2}$. Now suppose we let $\epsilon = .01$ and $\eta = .001$; then it can be shown that if we make about 30,000 throws, the chances are only 1 in 1000 (η) that the frequency of heads will vary from $\frac{1}{2}$ by more than .01 (ϵ).

Notice that Bernoulli's Theorem talks about the true probability rather than relative frequency after a finite number of trials. In applications, we do not know the *true* probability, but this theorem guarantees that as we increase the number of cases, we have greater and greater assurance that the relative frequency is becoming nearer

[1] Bernoulli's Theorem is one form of the so-called Law of Large Numbers.

and nearer the true probability. Thus, let a/n be the relative frequency at any given time, as above; this is in general calculable, since a and n are finite. Now let p be the true probability, which is unknown. Let P be the probability that a/n varies from p by less than a certain fixed amount, say .01; i.e., P is the probability that $| p - a/n | < .01$ (where $| x |$ is the "absolute value" of x, i.e., $= x$ if x is positive, and $= - x$ if x is negative).

Then Bernoulli's Theorem tells us that P approaches 1 (i.e., certainty) as n becomes larger and larger; hence, as we increase the number of cases we become more and more certain that the true probability lies within a specific range.

This, then, is an answer to the first objection raised concerning the definition of the true probability in terms of the relative frequency. We can never be *certain* that the relative frequency is approaching the true probability, but we can say that the more trials we take the greater becomes the probability that such is the case.

A simple example of the frequency theory in application should help us to make clear what is still obscure in the method. Mendel's experiments in plant hybridization, conducted in 1866, afford as good an example as any, since they employed the frequency theory in a rather simple form (6).

The experimental object was to trace the development of certain plant characteristics in hybrids; in this case the plant chosen was the pea. Mendel defined a "dominant characteristic" of a plant to be one which appears more frequently than not in the future generations of hybrids, and the chief problem of the experiment is to determine the relative frequency of the dominant characteristic in certain generations. Mendel chose many characteristics to observe, but it will be sufficient here to talk about only one, the characteristic of "round" seeds as opposed to "angular" ones. The dominant characteristic was "round." The first generation occurring from a cross between the round and the angular type proved uninteresting, since the dominant characteristic appeared almost invariably. However, in the second generation, variations did occur. We may map these out in a table, showing the increase in the number of plants used, the number of round seeds occurring

(the subclass), the number of nonround or angular seeds, the consequent reference class (the addition of round and angular seeds), and finally the relative frequency of the round seeds, calculated by dividing the total number of seeds into the number of round ones:

Number of plants	*Number of round* $= a$	*Angular*	*Total number of seeds* $= n$	*Relative frequency of round* $= a/n$ [1]
1	45	12	57	.79
2	72	20	92	.78
3	96	27	123	.78
4	115	37	152	.76
5	147	48	195	.75
6	173	54	227	.76
7	261	78	339	.77
8	283	98	381	.74
9	311	104	415	$.75^-$
10	336	111	447	$.75^+$

Thus as n increased, the relative frequency of round seemed to approach $\frac{3}{4}$; this result also occurred when other dominant characteristics were tested (in the end Mendel had 19,660 cases in the second generation, where 14,949 were dominant, or a relative frequency of $.760^+$ was obtained), and this led Mendel to formulate his famous law to the effect that in the second generation of hybrids the dominant characteristics have a probability of $\frac{3}{4}$ of occurring. The experiments were carried to further generations and other frequencies determined.

An early example of an attempt to measure the probability of a certain causal relationship is due to Thomas Bayes. Bayes, in effect, proposed the following problem: "What is the probability, given a certain event X, that it was caused by some other event, A_1?" Suppose now we know that an event occurs with a certain relative frequency when A_1 occurs, and with a certain relative frequency when A_2 occurs, etc.; suppose also that we have a case of X's occurrence. Then, Bayes asks, what is the likelihood that A_1 occurred before X did, or in terms of causality what is the likelihood that A_1 caused X? Bayes' Theorem is as follows: Let $A_1, A_2, \cdots, A_n$ represent all the possible causes of X. We wish to determine the probability that A_1 caused X. First determine by the usual methods of the frequency theory the following probabilities:

[1] Calculated to nearest second decimal.

p_1 = probability that A_1 occurs
p_2 = probability that A_2 occurs, etc.

q_1 = probability that when A_1 occurs, X also occurs
q_2 = the probability that when A_2 occurs, X also occurs, etc.

Then the probability that A_1 caused X is given by the fraction

$$\frac{p_1 \times q_1}{\sum_{i=1}^{n}(p_i \times q_i)}$$

where

$$\sum_{i=1}^{n}(p_i \times q_i)$$

means $p_1q_1 + p_2q_2 + p_3q_3 + \cdots + p_nq_n$.

Thus, for example, if three people are firing at a target, and a shot goes into the bull's eye, we calculate the probability that A_1 fired the shot as follows. We find that p_1, the relative frequency of A_1's shots, is $\frac{1}{2}$; that p_2, the relative frequency of A_2's shots, is $\frac{3}{8}$; and p_3, the relative frequency of A_3's shots, is $\frac{1}{8}$ (i.e., A_1 fired half the shots, A_2 $\frac{3}{8}$ of the shots, A_3 only $\frac{1}{8}$ of the shots). (Note that $p_1+p_2+\cdots+p_n$ must equal 1.) We also find that A_1 hits the bull's eye with a frequency of $\frac{1}{4}$, A_2 with a frequency of $\frac{1}{3}$, A_3 with a frequency of $\frac{1}{2}$. (Note that $q_1 + q_2 + \cdots + q_n$ need not add up to 1.) Then the probability that A_1 fired the given shot is

$$\frac{\frac{1}{2} \times \frac{1}{4}}{(\frac{1}{2} \times \frac{1}{4} + \frac{3}{8} \times \frac{1}{3} + \frac{1}{8} \times \frac{1}{2})} = \frac{2}{5}.$$

The application of Bayes' Theorem is very difficult, since, among other things, the theorem presupposes that we know all the possible causes of a given event, and the probabilities of their occurrence. In fact, the application of Bayes' Theorem seems to demand all the general presuppositions statistical empiricism is designed to avoid. Consequently, within the Frequency Theory it has become necessary to use other methods for measuring our degree of confidence in a causal relationship.

Before proceeding to a critical examination of statistical empiricism, and the Frequency Theory in general, let us review the fundamental tenets of the position:

1) The meaning of any scientific question must be reducible to certain ultimate data of sensation.
2) Consequently, the meaning of all general hypotheses about the natural world must be reducible to such directly confirmable sentences.
3) The *relative frequency* of an event is an objective measure of the scientist's degree of confidence in its appearance under certain specific conditions; it is a measure which is reducible to directly confirmable sentences, and the measure is a sufficient answer in general to questions regarding the laws of nature.
4) The relative frequency is an estimate of the true probability of the event, and with increase in sample size the estimate will approach the true probability "stochastically," in the sense of Bernoulli's Theorem.

Criticisms of *this* formulation of the Frequency Theory should be levelled at each of its points. We shall eventually want to show that "directly confirmable sentences" do not exist as a distinct class within science, even in a relativistic sense. For the present, several comments on the operational and positivistic definitions of concepts will be pertinent.

The opponents of positivism have found their task no difficult one. If all meaningfulness is to depend upon directly confirmable propositions, then what about the perfectly general question: which propositions (or sentences) are directly confirmable? Can this question be answered by reference to direct sensation? Does it really appear that the positivist has left us much better off than the mystic with respect to the solution of the problem of meaning? The mystic asserts that true meaning is to be found in certain experiences whose content cannot be translated into language. The positivist finds meaning in a series of propositions the confirmation of which cannot be a part of scientific method. Is it any better to assert that "*a* is red" is an obviously confirmed statement, than to assert with Leibnitz, say, that all existent particles are monads endowed with a purpose?

Operationalism also in the final analysis seems to fare no better. We cannot help appreciating the motives that drove Bridgman to

assert his positivistic thesis in physics. Physics, for all its experimental "front," had continuously been troubled with nonexperimental concepts that confused the statement of the theories, the arguments for their validity, and the inferences drawn from them. Bridgman's thesis is simply that whatever the nature of a concept, we must demand that the operations be such that they can be unambiguously and straightforwardly performed (1).

Although this description serves a very useful purpose in warning the scientist about the risks entailed in divorcing concepts from experience, as an analysis of scientific method it fails at the most crucial point. The analyst of method must be a generalizer, and his object should be a general description of the conditions under which inquiry takes place. In this connection it does not help merely to affirm that there are "operations about whose performance we have no hesitation." The obvious but specialized objection to such statements is that a lack of hesitation in these matters often indicates ignorance; for to the uninitiated in laboratory techniques many operations look "simple and obvious." The generalized objection is that without at least an approach to defining the class of "unambiguous operations" we have failed to make the most important step of the analysis.

An interesting application of the operationalist viewpoint on the problem of confirming empirically the presupposition of randomness is made by W. A. Shewhart (10).

Randomness can be defined within the realm of *formal* theory in a comparatively simple manner, as we have done in Chapter II. But though we may be able to attain an exact definition in formal probability theory, it is not easy to see how randomness can be so defined that we can confirm our belief that a set of observations are really representative of a universe. And such an empirical confirmation does appear necessary, for the legitimacy of application of statistical theory depends upon an assumption of randomness of selection. Whether we are making measurements of the distance between two astronomical bodies, or counting the harvest in certain fields, we usually assume a randomness of observation. The rationalist solution of the application of randomness, based on the assumption of equiprobability, is well known to lead to awkward paradoxes,

and represents an intuitive outlook of a very restricted character. For one thing, "no evidence" is itself a difficult matter to establish; for another, if ignorance is made a sufficient condition for establishing a certain hypothesis, then it is little wonder that contradictory hypotheses can be derived.

Shewhart proposes as a solution to the problem of randomness the thesis that the "only operationally verifiable way to define random order is in terms of some chosen random operations." While it follows that there is no unique description of randomness, it does seem to follow that for any operational definition an exact procedure of test can be instituted. The operational model Shewhart chooses is one in which an observed sequence is drawn one at a time with replacement and thorough mixing by someone who is blindfolded (10).

One is certainly led to doubt, from Shewhart's account of this example of operational definition of randomness, whether the demands of freedom from ambiguity are in any sense satisfied. It is an obvious enough criticism to point out that "replacement" and "thorough mixing" are ill-defined concepts within our language, but the real point of criticism of the operationalist definition lies somewhat deeper. How do we know that a certain sequence of operations "defines" a concept? Is the selection of operations completely arbitrary, or is there a difference within science between a "correct" and an "incorrect" definition? Would it not be naive to assert that *any* operational definition of mass, or life or randomness constitutes a legitimate definition provided one uses it consistently? On such a basis, science would be replete with concepts used in as many ways as there are individual workers. No, the scientist's definition of concepts must be conditioned by the history of the science, and by the work of his contemporaries. It would be simply incorrect to define a random sequence as one arising from the selection of every third item in an ordered set of objects. Such a definition, for all its operational precision, would conflict with present practice within experimental science. Consequently, the adequate definition of a concept is a problem not arbitrarily decided in science, and the analysis of the method of defining concepts is an important aspect of scientific methodology. (See Chapter XI.)

Returning now to the Frequency Theory, we note its validity depends upon the truth of the assertion that the fraction a/n can be said to have a limit as n approaches infinity. In what way is Bernoulli's Theorem "confirmed" in nature? Can we construct an empirical proof of its confirmation? Now it is true that Bernoulli's Theorem follows deductively in a mathematical system of probability; i.e., it can be proved if we grant certain assumptions within a formal theory of probability. But the presupposition of the Frequency Theory is that the relative frequency is an estimate of the true probability, and hence it must be shown that the theorem really "works" in actual experience. Naively, one must expect that the theorem could be tested in experience by throwing a coin a great many times to see whether, in fact, the relative frequency of heads to tails approaches $\frac{1}{2}$ as the number of throws grows larger. It should be clear to the reader from our previous discussion that such experiments are futile. Even though the relative frequency of heads to tails did approach $\frac{1}{2}$, this would hardly verify Bernoulli's Theorem unless we knew beforehand that the true probability of a head's appearing *was* $\frac{1}{2}$. Any results from repeated trials of the sort described would have to confirm Bernoulli's Theorem, since this theorem is used to calculate the true probability from the relative frequency obtained. Actually, nature presents us with many illustrations of sequences of events that do *not* approach any limit; the relative frequency of deaths of persons over sixty has not approached a certain unique limit over the years. In this case, we say that the reference class has "changed." Consequently, we say that the relative frequency approaches a unique limit provided the reference class remains constant. But the only test of such constancy must lie in the observed relative frequencies, and hence the empirical test of Bernoulli's Theorem depends upon whether or not these frequencies approach some constant limit. As a result, we must assert that (1) the adequacy of relative frequencies as estimates of the true probability depends upon the truth of Bernoulli's Theorem, while (2) the truth of Bernoulli's Theorem depends upon the adequacy of relative frequencies as criteria of constancy of the reference class.

This circularity, inherent in the formulation of the Frequency

Theory we have given, can be removed by a reformulation of the theory in accordance with the methods discussed in Chapter II.

The revised account of the Frequency Theory runs as follows: Let H_0 be a certain hypothesis about the natural world; let H_t be a set (not necessarily finite) of alternative hypotheses and let a set of data be collected. Then to test whether H_0 is to be accepted or not, we derive, assuming H_0 to be true, the relative frequency with which the observed data (or better, a mathematical function of the observed data) would occur if H_0 holds. If this relative frequency is small, then we abandon H_0 in favor of one of the alternatives. This description is quite general, and includes the more restricted formulation of the frequency theory which usually reads somewhat as follows: in order to determine our confidence in a given hypothesis about nature, we determine the relative frequency of successes of the hypothesis, and the nearer this approaches unity, the greater our confidence. Further, the methods of Chapter II show that, contrary to the writings of many empiricists today, e.g., (9), the measure of our confidence in a theory is not expressible as a unique number. We have first to define a method of selection, and in terms of this method compute the possible "risks" involved in accepting any alternative. Our confidence in a hypothesis is thus a function of what we take to be alternative hypotheses, and the concomitant risks involved in rejecting them. Thus we might have a high degree of confidence in Newton's Law of Universal Gravitation if the only alternative hypothesis were that bodies of equal mass attract each other inversely as their distance squared. But if a possible alternative is that bodies of equal mass attraet each other inversely as the square of their distance *plus* a constant of the order 10^{-50}, our confidence in Newton's Law would be much less on the basis of present physical data.

The revised account of the Frequency Theory therefore asserts that the methods of Chapter II are general, *except* for certain questions regarding the formal probability theory, and except for certain questions concerning the data of sensation.

In general, empiricists regard the formulation of theoretical sciences to lie "outside" the realm of empirical methodology, in the sense that the data of experience are not necessary conditions for an

adequate formalism. Thus, within purely formal science the historical empiricists were actually rationalists (in the sense explained in Chapter V). Consequently, the empiricists' solution of the choice of a fundamental probability theory has not differed from that of the rationalists' we have already discussed.

The empirical attitude towards observations should be clear from the analysis we have already given. In terms of the methodology presented in Chapter II, he asserts that there are some questions of fact the answering of which does not involve any risk. These basic questions then form the origin of scientific inquiry, and in terms of them we are to develop the answers to all the other questions of methodology raised in Chapter III. In this sense, statistical method is not completely general, since it is not applicable to the simple data of sensation and reflection. Thus for the empiricist, when I decide that I now see a "red patch," I make but one "observation," and the alternative hypothesis is rejected without any concomitant risk.

Now not all questions of fact receive such direct confirmation. If I ask whether there are mountains on the moon, the empiricists can answer affirmatively without risk by using observations through a telescope. But if I ask whether there are mountains on the *other side* of the moon, I can only answer with a certain degree of confidence, based upon the distribution of the directly observable mountains on *this* side. Thus empiricists divide the facts of science into two classes: the directly observable and the indirectly observable. To make their analysis of scientific method complete, they must supply some criterion for distinguishing between these two classes. The discussion of the subsequent chapters will attempt to show the futility of a search for such a criterion.

The empirical attitude towards the techniques of evaluating methods of selection are not generally discussed in the literature, since this phase of methodology is not generally known. However, one would expect here the typical empirical solution of value problems. The best method of selection depends upon the kind of "losses" we are willing to incur. Now traditional empiricism has always measured (ultimate) profits and losses in terms of *utility*, either for an individual or a social group. Utility itself is measurable

in terms of pleasure, which, according to Bentham at least, is representable on a one-dimensional scale. Consequently, the simple facts which form the basis of value measurements are the feelings of pleasure (or pain) we have as a result of some impact of the world upon our senses. These feelings are data of our internal sense, and for the empiricist they are immediate; i.e., we incur no risk in deciding whether we do or do not feel pain or pleasure. Now a thorough methodological analysis of the meaning of pain and pleasure judgments, and of sensations in general, should be sufficient to show that these judgments cannot be immediate in this sense. It is to such an analysis that we turn in the next chapters.

REFERENCES

1. Bridgman, P. W., "Operational Analysis," *Philosophy of Science* (1938).
2. Carnap, R., "Testability and Meaning," *Philosophy of Science* (1936–37).
3. Carneades, as reported in E. Zeller's *History of Ancient Philosophy*, Harcourt (1931).
4. Hume, D., "Enquiry into Human Understanding," in *English Philosophers from Bacon to Mill*, edited by E. A. Burtt, The Modern Library (1939).
5. Hume, D., *A Treatise on Human Nature*, ed. by L. A. Selby-Bigge, Oxford University Press (1894).
6. Mendel, G., "Experiments in Plant Hybridization," translated by W. Bateson, in *Mendel's Principles of Heredity*, Cambridge Univ. Press (1909).
7. Mises, R. von, *Probability, Statistics, and Truth*, Macmillan (1939).
8. Pearson, K., *The Grammar of Science*, London, Black (1900).
9. Reichenbach, R., *Experience and Prediction*, Univ. of Chicago Press (1938).
10. Shewhart, W. A., *Statistical Method from the Viewpoint of Quality Control*, Graduate School, Dept. of Agriculture (1939).

Chapter VIII Criticism

The schools of rationalism and empiricism constituted an "antinomy," a conflict of ideas, that in the history of modern epistemology threatened to leave philosophy forever stranded on the shores of scepticism. On the one hand the rationalist method was supposed to enable one to construct a definite, deterministic picture of nature, such that by an application of the rigorous methods of formal logic one could answer questions about the past, present, and future, and could answer these questions without the possibility of error. But in the end, the methodology of formal science turned out to be incomplete in itself; even the science of language-analysis cannot be made to answer all its problems within its own framework. On the other hand, empiricism laid claims to an ability to discover the particular facts about the world, but was forced into ambiguous criteria concerning the meaning of the immediate data of sensation. The solution of Immanuel Kant of this antinomy, designed to save the claims of both the rationalist and the empiricist, was based on a further analysis of these facts that form the starting point of all knowledge.

And, indeed, as our study has progressed it has become more and more apparent that the general solution of the problem of truth cannot be given until we have further analyzed the process by which the inquiring mind answers questions of fact. For the empiricist there are at least some questions of fact whose answering requires no analysis, but if the method of science is to be precise and general, it becomes its task to define or characterize these immediacies. It was the thesis of Kantian philosophy that if such immediacies do exist, they do not have the power of coordinating themselves into

any meaningful structure and that the accomplishment of such coordination by the mind necessarily presupposes certain laws that could not by their very nature depend on facts for their answering. It is not our purpose here to follow the intricacies of Kant's argument in his *Critique of Pure Reason* (1781); our aim will be better served by analyzing a rather simple experiment in order to discover how the experimenter "constructed" his result.

We choose Galileo's study of the inclined plane, the object of which was to calculate the motion of a ball rolling (unmolested) down an incline. Marks were made on the incline, indicating the distance traveled. Not having accurate timepieces at his disposal, Galileo resorted to a rather ingenious device: he took a tank of water so constructed that the flow of water from the pan could be controlled by placing a finger over an outlet. When the ball started to roll, the finger was removed, and when the ball reached the required destination, the flow was stopped. The amount of water released during this time was preserved in a jar, and now the times taken by the ball to traverse various distances could be measured by the corresponding weights of the water in the jars, these weights being calculated rather precisely by means of the relatively excellent scales of Galileo's time.

Galileo's laws of the inclined plane were "induced" from a set of facts of the sort "The ball took x seconds to travel y inches." It will be evident enough that none of these facts were immediate in the sense that their determination presupposed no other knowledge. Among other things, in answering the question as to how long the ball took to travel a given distance, Galileo presupposed: (a) certain geometrical principles in placing the marks on the incline; (b) that the flow of water from the tank was constant, i.e., that the same amount of water flowed out of the tank during a given period regardless of when the experiment was conducted; (c) that the observer's "reaction time" in placing his finger over the outlet of the tank when he saw the ball pass a certain mark was constant.

Certain of these presuppositions might be taken as "definitionally" true in the operationalist's sense; e.g., the assumption regarding the rate of flow of the water being proportional to time might be simply a definition of time in Galileo's sense, an operational defini-

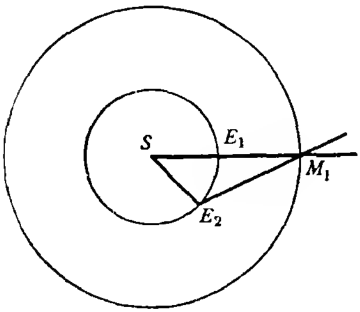
tion being no more than a method of answering questions. But some of the assumptions do not appear to be verified in so facile a manner, and, indeed, their denial would make significant changes in the manner in which the questions were answered (e.g., if we deny that the reaction time of the observer is independent of the result, we would obtain, presumably, different answers to such questions as "When did the ball reach *B*?").

Another famous example will help make clear the point of view we are developing. When Kepler discovered his famous laws of planetary motion, he employed the results of his predecessor Tycho Brahe; to the reflective empiricist, the determination of the law of motion of a planet seemed to be as follows:

> The object of Kepler was to determine the real path described by each of the planets, or let us say by the planet Mars (since it was of that body that he first established the two of his three laws which did not require a comparison of planets). To do this there was no other mode than that of direct observation: and all which observation could do was to ascertain a great number of the successive places of the planet; or rather, of its apparent places. That the planet occupied successively all these positions, or at all events, positions which produced the same impressions on the eye, and that it passed from one of these to another insensibly, and without any apparent breach of continuity; thus much the senses, with the aid of the proper instruments, could ascertain. What Kepler did more than this was to find what sort of a curve these different points would make, supposing them to be all joined together. He expressed the whole series of the observed places of Mars by what Dr. Whewell calls the general conception of an ellipse (3, pp. 338–339).

Now Mill omits a description of the manner in which Kepler determined the various positions of Mars; as would be expected, the actual account of Kepler's work shows that these were not determined "immediately," without any presuppositions. Neglecting certain factors, we may briefly describe the locating of Mars on its path as follows. The first observation occurs when the positions of the Earth, Mars, and the Sun lie approximately on the same straight lines. To determine completely the position of Mars relative to the Sun and Earth, we must find another line on which the planet falls and take the intersecting points. This second line is determined by allowing Mars to complete its orbit (686.95 days); the position of the

Earth at the end of this time can be calculated if we know the Earth's period (approximately 365.25 days) and that the Earth's orbit is approximately circular; by determining the angle between the Sun, Earth, and Mars, we can plot another line on which Mars must fall. A simple diagram is appended.



The relative position of Mars is determined by getting the points SE_1M_1 on a straight line (at position noted by E_1 and M_1); then a period of 686.95 days having passed, M has resumed its original position, while E is in the position E_2; we determine the angle SE_2M_1, and locate M_1 on the intersection of SE_1 and E_2M_1. The *absolute* position of Mars (in terms, say, of the Earth's diameter) is calculable if we know SE_2, the distance of the Earth from the Sun when the Earth is at E_2.

Neglecting the presuppositions necessary for taking angular measurements, Kepler's determination of one position of Mars presupposes: (a) that the period of Mars is 686 95 days and that Mars returns to its original position in the given time, (b) that the Earth's orbit is (approximately) circular, (c) the period of the Earth. Thus, to make even one measurement of Mars' position, general laws of astronomy had to be presupposed, *including laws of planetary motion.*

If we pass to modern examples of experimental procedure, it will be still more evident that the facts the experimenter collects presuppose many principles for their determination; the complicated timepieces of modern experiments depend on a whole set of kinematical and mechanical principles.

One might now argue that the principles presupposed by the experimenter are verified by other facts. If this thesis is correct, it becomes very evident that in order to determine the nature of experimental method it is necessary to determine what are those questions of fact that presuppose no laws or other facts in their answering.

We note first that all questions requiring a quantitative answer

(i.e., a number of some sort) are not questions receiving an immediate answer. For to measure anything, an instrument of measurement is required, and all such instruments presuppose the principles by means of which they were constructed; the ruler presupposes the geometrical laws of congruent segments and bisected segments, the clock presupposes mechanical laws of springs, and so on. Even discrete counting presupposes laws of addition and certain principles of succession. One quantitative question might be considered as immediately answerable, namely, the question demanding "one" as an answer; if we are to answer how many objects are on the table, and only one object lies there, apparently we can give our answer without reflection or presupposition. But a moment's thought will soon show us that even such a simple question cannot be answered unless we have some criterion for judging what constitutes unity; one person might consider a book to be a single object, another might consider it to be a great many objects (the pages) bound together. Whatever our criteria may be that allow us to say that a thing is a unified whole, they are laws, and they are laws that must be presupposed by the inquiring mind.

If we accept the conclusion that quantitative questions are not immediately answerable, then we must also acknowledge that what immediacies there are must be nonquantitative, i.e., "qualitative." That not all qualitative questions are answerable immediately is again rather apparent. I cannot say whether this object is a white piece of paper until I have determined whether it is paper or not, and such determination cannot be made immediately; no one would profess to be able to decide without presuppositions the texture of a given object.

Again, from the Kantian point of view, those questions demanding qualitative *relationships* between objects for their answers (e.g., "Is this book on top of that one?") are not answerable immediately; the given relationship must be explained, and its explanation consists of a set of criteria by means of which we judge whether it applies. If I am to answer, for example, whether this book lies on top of that one, I must take the two sensations of the two objects and compare them by the criteria of "on-ness" to determine the truth or falsity of the question asked. The argument may be phrased

as follows: (1) All propositions asserting a relationship between things require at least two *relata*. (Thus, in the proposition "*A* is the brother of *B*," *A* and *B* are the relata of the relationship "is the brother of.") (2) To determine the answer to questions concerning the relationship of two objects to one another (e.g., "Is *A* next to *B*?") we must answer certain questions regarding the relata (e.g., "Where is *A*?", "where is *B*?"). Hence (3) questions concerning relationships cannot be answered immediately since they presuppose the answering of other questions.

The propositions about qualities that remain as candidates for "immediacies" seem (at best) to be ones declaring that we have a sensation of a color, or a pain, or a taste, or a smell, etc., in our consciousness. But even such judgments, in that they relate a sensation to consciousness, are not known without presupposition, since a *relationship* between objects is not discoverable immediately; in the end, the immediacies must simply be isolated bits of "intuitions" (*Anschauungen*) made without reference to anything else. The "given" in sensation is so elementary that no one can really reflect on it without distorting it. Even the naming of the colors, pains, smells, etc., that are immediately observable is a distortion. If I am to decide whether I now have a sensation of red, I must also decide whether the sensation I now have is similar to another one I have previously termed red, and this decision, like all memory decisions, cannot be made immediately, since it is one of comparison; the question requires us to determine the relationship between two sensations, and hence by our previous argument cannot be immediate. Immediacies, then, are so private a matter that not only can no one else but the observer decide their truth, but the observer himself cannot name them or explain them adequately. They have now turned out to be discrete, unrelated, inexplicable sensations.

Such at least the proponents of criticism take them to be, and in the light of this analysis they raise a very significant question: how does the mind come to answer questions concerning relationships and quantities in terms of these immediate sensations alone? The empiricist had demanded of science that it construct its view of the world only in terms of what was immediately known in sensation and reflection. But for the criticist this demand is impossible to fulfill,

for the very nature of these immediacies precludes the possibility of our constructing by means of them alone a world of relationships and laws. No, the "hard facts" of Locke are not sufficient to answer any other questions; starting with the immediacies of sensation and reflection *alone*, we end up with these immediacies and nothing more. To gain more, to construct a world from observed facts, we must presuppose laws at the very start, and they must therefore be laws that are not induced from observed data. When these laws are applied to the immediacies of sensation and reflection they allow us to construct the world of science, i.e., to answer other questions. These laws Kant called a priori to distinguish them from the a posteriori laws later induced from observed data, and *synthetic* to distinguish them from the *analytic* laws, the truth of which is known by definition.[1]

In Kant's words, the mere "given" of sensation, without the aid of a priori laws, would leave us like animals, intellectually blind. To have thoughts, i.e., to make *judgments* or propositions having both a subject and predicate, we require equipment. It was the task of the main section of the *Critique of Pure Reason* to determine the necessary conditions for constructing such judgments.

The Kantian problem can best be understood with reference to the discussion of method in Chapter II. In order to respond to certain questions in terms of observations, it is necessary to know something about the universe from which the observations are drawn, and something about the manner in which they are drawn. In general, the techniques discussed in Chapter II make a universal demand for randomness of certain groups of observations. In this sense, randomness is an a priori condition for the application of statistical methodology. Now one might argue, as the empiricists have done, that the assumption of randomness itself may be tested by the observations, and that in this sense randomness is a posteriori. But an examination of all prevailing techniques of "testing" randomness shows that the presupposition is re-employed in another form. The operational definition of randomness, characterized by a

[1] Our analysis of rationalism has already shown the futility of making an *absolute* distinction between the analytic and synthetic, but this analysis was the result of viewpoints more recent than Kant's day.

set of physical behavior patterns, evidently does not imply an absolutely specific pattern. Such an implication would mean that in practice we could only guarantee randomness of observation where our behavior met absolutely rigid specifications based on a mechanical description. We would therefore have to assert that in effect randomness was never attainable. Instead, the operational definitions represent a "morphological" classification of behavior patterns. If a behavior pattern belongs to a certain type, i.e., is representative of a class of rigid patterns, then it conforms to the required definition. Now we cannot guarantee that a behavior pattern will have any specified characteristic; we simply suppose that the elements in nature beyond our control will not operate to the extent of producing erratic behavior. Hence, in order to assert that a certain set of observations were made "at random," according to operational criteria, we must first *presuppose* that the physical behavior which produced the observations is representative of a class of behavior patterns. But the criteria of representativeness are exactly those of randomness; that is, to assume that a behavior pattern is representative of a class of patterns is to suppose that the pattern has been selected at random from a class of patterns, just as the assumption that a chip drawn from a bowl is representative of the bowl's contents is equivalent to the assumption that the chip has been drawn at random.

The conclusion of the argument is therefore quite simple: *We would not be able to find randomness in our observations had we not first put it there in some form.*

It is true that within statistical literature one finds objective tests of randomness within a set of observations. An examination of such tests, however, will show that the presupposition of randomness always appears in some form with respect to a subset of the observations (e.g., that the single observations, or certain groups of the observations, have been drawn at random).

The point of view of criticism is now clear enough: *the questions which pure sense data answer are not rich enough to provide us with the presuppositions we must make in order to answer all questions within science.* Empirical data alone can never even begin to justify the mathematical and physical constructs that we must assume in

employing the methodology of Chapter II. Empiricism simply fails to show how any such methodology could ever be justified on the basis of sense data, for the meaning contained in a datum of sensation is far too trivial.

We have chosen randomness as an example of the critical argument, and needless to say, the example is not Kant's own. Kant's general problem was the following: to determine what the learning mind must put into experience in order to give objective answers to questions about the natural world. We shall consider two of Kant's suggestions: the demand for principles of individuation, and the demand for principles of regularity. From the point of view of this essay, Kant's original formulation must be translated into terminology more useful for the modern theory of inference. The thesis is the same as that already given: if observations are to be used in the capacity demanded of them by the methods of Chapter II, then what general presuppositions must we supply over and above what is immediately given in sensation? That is, "observation," in the sense of statistical methods, cannot mean only "immediate sense data"; we have now to determine what else is implied in the term.

1. The problem of *individuation*, considered as a problem of scientific method, is concerned with the method of answering factual questions of the sort, "Is x the same as y?" Evidently such questions are rather easily answered in most cases as long as we have a way of comparing the properties of things. This table is not the same object as that chair, for they have some rather obvious differences; my observations tell me that one is smaller than the other, for example. Now any *general* method of individuation must suffice to answer *all* questions of this type. Is the method of description sufficient for this purpose? Leibnitz, among others, seems to have thought so, and claimed that the basic method of differentiation of two objects was by means of "minute description," with the consequence that if two objects were identical in all their attributes, they could not be two distinct things, but must be one and the same. But we may reflect, along with Kant and others, that not only is there no guarantee that the method of minute description may not lead to contradictions, but, what is more to the point, that the

method really presupposes that individuation has already taken place; to make a complete comparison between x and y, I must already have considered x and y as objects, and nothing can be considered as an object of experience unless it has become an individual, i.e., been individuated. Kant found two conditions to be necessary and sufficient for individuating all objects of experience, namely, that they be located in space and in time.[1]

Only one object can occupy a given place at a given time, so that space and time are all the conditions necessary for individuation. When Kant's solution is translated into the language of experimental science, it means that no experimenter can undertake a series of observations without first having placed the objects to be observed in a space-time framework. And since experience supplies directly only discrete, unrelated sensations, space and time cannot be derived from the sensations we have had. They are the necessary conditions for a meaningful experience; they are a priori, and the nature of the space-time system used to individuate objects does not presuppose the knowledge of fact.

The notion that space and time are not objects of immediate sensation is a hard one to grasp, since one is apt to think that our observation of extension is as immediate as our observation of some color or pain. It is worth while, then, to pause a moment in the argument to explain what sort of space and time we have in mind. We say that a point is *individuated* in (3-dimensional) space if there exists a set of "coordinate axes," i.e., 3 non-coplanar intersecting lines, and a method of measuring the distance of the point from these lines. In the classic (Cartesian) analytic geometry the lines were perpendicular one to another, but they need not be. In some experiments, a two- or even a one-dimensional coordinate system is sufficient; for example, a point on this page may be individuated from all other points on the page by measuring its distance from the top and the left side; in Galileo's experiments on the inclined plane, a one-dimensional space sufficed, the position of the ball being given by its distance from the starting point. Evidently, space alone is

[1] As an historical note of clarification, it should be pointed out that Leibnitz's "complete description" did not include space-time description, and hence did not include Kant's method as a special case, for space and time were not essential properties of the "monads" of the Leibnitzian world.

not sufficient for complete individuation, and a time-coordinate system is also necessary. It follows that the space-time framework presupposed by the experimenter is a *relational* one; it enables him to compare or relate two or more objects with one another; thus Galileo was able to relate the various space-time positions of the ball with one another. Hence, it is not the extensional attribute of space, but its relational attribute that makes it a priori and independent of immediate observation; for the immediacies of sensation, whatever they may be, are not relations. The reader may rightly wish to raise an objection to certain expressions in the explanation above; we have said that a point may be individuated by measuring its distance from certain lines, but clearly such measurement implies that we have already considered the point as an individual. Rather, we should say that anything having a certain specified distance from the coordinate axis *is* a point and *is* individuated from all other points. More precisely, space and time are terms used to represent all the presuppositions we must make in individuating an object for the purpose of observation.

Space and time also allow us to coordinate our immediate sensations into some meaningful pattern; once our various experiences are placed in some sort of space-time framework, we can compare them and construct objects having a number of observed properties. For example, I can construct the idea of this table with its legs and top by means of a space framework; placing my various sensations in different times allows me to distinguish them and give a successive description of an object. As a consequence, observations need not be single entities but may take the form of pairs of elements, or n-tuples, of the form "x occurred at time t_1 at coordinates y_1, y_2, y_3."

2. But space and time are not enough to permit us to construct an understandable world; or rather, the conditions under which a space-time framework can be used by the experimenter should be made clearer. We need only consider how the experimenter differentiates objects in time. He does this by means of a timepiece which must be so constructed that changes occur in it at regular intervals; that is, the timepiece is constructed to obey some "mechanical" law; without some such regularly operating mechanism

to rely on, the experimenter would have no way of calculating time. Questions about time would still be meaningful, perhaps, if all clocks stopped, but such questions would be meaningless if there were no regularity in nature at all, for then there would be no way of determining the passage of time except by a mystical intuition of duration à la Bergson, an intuition that is meaningless for the experimental scientist since it is inexpressible. If it is true that the answers to questions about time presuppose a regularity in nature, and the answers to time-questions are necessary conditions for answering all questions but the immediacies, then it follows that any experimenter interested in inquiring into the nature of the world must presuppose a regularity in that world. That is, the meaning of an observation presupposes a principle of regularity in nature. And here we have Kant's answer to the scepticism of Hume: the world is regular in its behavior, and we must expect it to behave tomorrow as it did today, simply because without such a presupposition all questions that science can raise (other than the trivial immediacies) become meaningless. Causal determinism in nature becomes the *sine qua non* of investigations into matters of fact and law. The naivete of the empiricist lies in supposing that investigations into matters of fact involve little or no problem, if we can only gather our immediate facts and make complex facts therefrom without the aid of laws, then questions of law seem to become meaningless or at least superfluous. But an examination of the method of constructing a nature which is understandable and in which objects are relatable one to another shows that some laws must be presupposed; no amount of immediate data can either prove or disprove that the world is regular or causally determined in its nature, but unless the world is so conceived, it cannot exist for the inquiring mind.

The Kantian position is not easy to grasp; indeed, it is all too often misinterpreted. Kant's demand or postulate that there exist a determinism in nature is not scientific wishful thinking. He does not mean to assert that a nondeterministic world would be a discouraging one for the scientist interested in formulating a description of nature. His demand is not one to be confirmed or refuted by examining the world of all possible experiences. Rather, for Kant *no* investigations can be made, *no* world or experience can exist,

unless a regularity in nature is presupposed. There would be no laws of nature, no facts of nature (except the immediacies) unless we had already imposed on nature a certain form. It is not that the mind, in a Humian sense, "unconsciously" puts regularity into the world; it is rather that the very possibility of mind and an observable world require as a necessary condition a natural determinism. This seems to be the point of the famous section of the *Critique of Pure Reason*, the "Transcendental Deduction." *If* we grant the possibility of objective experience, *then* we must also grant (among other things) an "unbroken mechanism" in nature; the one is deducible from the other, and hence, in the logician's terms, the regularity of nature is the necessary condition for an investigation into nature (X being a "necessary condition" for Y, if Y implies X).

Thus Kant "synthesizes" the opposition of rationalism and empiricism by asserting that the knowledge of *both* law (the a priori) and fact (the immediacies) are necessary for any experimental method.

The formulation of the laws of spatial relations is called the science of geometry, those of temporal relations, kinematics, those of causal relations, mechanics (and possible physics). The sciences of geometry, kinematics, and mechanics all have an a priori aspect as does the science of arithmetic (number), since arithmetical laws are presupposed by geometry; for Kant, these a priori laws are not dependent on observation (immediate sensations) for their verification.

To summarize the pertinent aspects of Critical Philosophy:

1. In order to answer questions about the natural world one must make objective observations.

2. The following are necessary conditions for objective observations:

   a) an immediately given in sensation.
   b) presuppositions (a priori) of individuation (geometry and kinematics).
   c) presuppositions of regularity (mechanics).

(And to this list, we have added:

   d) presuppositions of randomness.)

The simplest type of experiment must make these presuppositions. For example, suppose we ask how heavy an object is at a certain moment of time. In order to make a single observation of the form "X was observed to weigh p lbs. at time t_1," we must have presupposed (explicitly or implicitly) principles of identification, principles of time pieces, principles of randomness. And these principles cannot be substantiated by the immediacies of sensation, since the intuitions of sensations lack all aspects of generality.

The philosophy of criticism as outlined by Kant did not analyze a problem that became pertinent after the gigantic mathematical developments of the nineteenth century; we have said that the a priori laws of arithmetic, geometry, kinematics, and mechanics do not depend on immediate sensations for their verification; on what, then, do they depend? That is, how do we investigate questions within geometry, kinematics, mechanics, etc.? This question had a rather obvious answer for Kant since he (along with the majority of scientists of his day) conceived of only one geometry (the Euclidean), one kinematics, and one mechanics (the Newtonian.)[1] Hence there seemed to be no difficulty in saying that the a priori laws are verified *intuitively*. But long before Kant wrote his *Critique*, mathematicians had been investigating the foundations of Euclidean geometry, preparing the way for the important results of Gauss, Bolyai, Lobatscheski, Riemann, etc., in the development of non-Euclidean geometries.[2]

These investigations suggested that many other non-Euclidean systems could be constructed, each denying one or more of the Euclidean postulates, and actually many of these have been developed. Further, Riemann's work suggested a more elaborate study, into geometries of more than three dimensions, a study that became an important branch of mathematics.

Not long after the discovery of non-Euclidean geometry, physicists developed a non-Newtonian kinematics. Kinematics is the science that, adding to the concepts of geometry the concept of

[1] Or, rather, a modified Newtonian Mechanics, since significant changes were made in the Newtonian theory (e g , by Boscovitch), though the fundamental postulates remained the same

[2] See Chapter V for a survey of the history of non-Euclidean geometry.

time, studies the velocity, acceleration, etc., of a point, or set of points, and describes the "behavior" of a moving point in an n-dimensional continuum. Kinematics, then, does not consider problems of mass, force, etc., but only the simple problems of the motions of points, groups of points, areas, and solids.

Before Kant's time there had been developed several conflicting systems of kinematics to describe the manner in which various points move in the universe. The Aristotelian theory of the circular movements of the heavenly bodies and the rectilinear movements of earthly objects, the Ptolemaic, Copernican, and Keplerian descriptions of the motion of the planets were all partial kinematical systems. But by the time the *Critique of Pure Reason* appeared (1781) the Newtonian kinematics had so far eclipsed or encompassed all others that there was little doubt in the physicist's minds that no other kinematics was feasible as an accurate account of the motion of particles.

A mathematical development of a non-Newtonian kinematics is made in Einstein's Special Theory of Relativity. This system denies one of the fundamental properties of Newtonian kinematics, the Law of Addition of Velocities, which states:

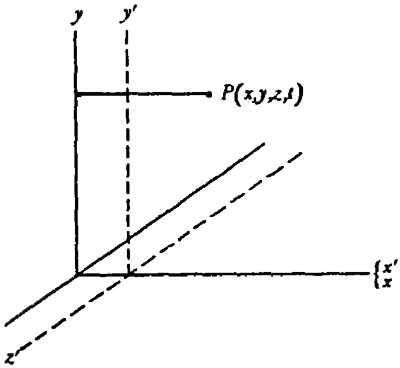
Let K be a system of points traveling at uniform velocity v in a straight line relative to another system K'; further let the point x travel in the same direction as K and with uniform velocity w relative to K; then x's velocity relative to K' is $w + v$.

For example, K might be a railway train moving along a straight track with a steady speed of 60 miles per hour relative to the earth (K'). If x is a man who starts down his car in the direction of the engine at the rate of 3 miles an hour relative to the train, then the Law of Addition of Velocities states that his speed relative to the earth is $60 + 3 = 63$.

This Law may be stated in terms of a coordinate system. We take three mutually perpendicular lines in the three-dimensional system K; then any point P of K will be determined by its distances from the three axes at a certain time, i.e., by the four numbers x, y, z, and t.

We suppose that at time $t_1 = 0$, these "axes" coincide with the axes of K'. The Law of Addition of Velocities allows us to deduce

that if x, y, z, and t are the numbers determining a point relative to K, then the numbers determining the same point relative to K', if K' moves at velocity v in the direction of the x-axis, are given by



$$x' = x - vt$$
$$y' = y$$
$$z' = z$$
$$t' = t;$$

that is, the distance from the y- and z-axes remains unchanged as does the time of P's arrival at this spacial point. These equations are called the "Galilean transformations."

In the Special Theory of Relativity, the Galilean transformations are denied, and hence also the Law of Addition of Velocities. Instead, this non-Newtonian kinematics gives the coordinates of the point in question relative to K' as follows (the Lorenz transformations):

Let c be a preassigned constant; in practice, c is taken as the velocity of light *in vacuo* (approximately 300,000 kilometers per sec.): then

$$x' = \frac{x - vt}{\sqrt{1 - \frac{v^2}{c^2}}}$$
$$y' = y$$
$$z' = z$$
$$t' = \frac{t - \left(\frac{v}{c^2}\right)x}{\sqrt{1 - \frac{v^2}{c^2}}}.$$

The consequences of the denial of the Galilean transformations are as astonishing as the consequences of the denial of Euclid's Parallel Postulate. But we should note that if v is small relative to c, then the Lorenz transformations reduce for all practical purposes to

the Galilean, for the important fraction v^2/c^2 is very small, and x' and t' reduce very nearly to $x - vt$ and t respectively. Thus in practice, even if v is as large as 30 km. per sec. (about 70,000 mi. per hour), v^2/c^2 is about 1/100,000,000, a quantity presumably less than the error of observation.

By rather elementary methods it can be shown that the Special Theory of Relativity has these consequences: to an observer at rest relative to K, segments that would appear equal if K' were also at rest relative to K, "contract" when K' is in uniform motion ("Lorenz contraction"), and (vice versa) to an observer on K' the segments on K contract, for if K' moves with velocity v relative to K, then K moves with velocity $-v$ relative to K'. Thus, to one standing on the railway bank, the measuring rods of the observer in the railway carriage appear smaller than those on the bank, though initially all rods had an equal length. Next, to the observer on the bank, the velocity of a body moving with velocity w relative to and in the same direction as K' is not $w + v$, but $w + v/(1 + wv/c^2)$, i.e., slightly slower than $w + v$ if w and v are small relative to c. The speed of light is constant relative to all moving systems, for the speed of a light beam in the railway car to the observer on the bank would be given by the above formula, where $w = c$:

$$w' = \frac{c + v}{1 + \frac{cv}{c^2}} = \frac{c + v}{1 + \frac{v}{c}} = c.$$

Finally, no system can travel faster than c relative to another, for then all its segments would contract to minus values.[1]

Other non-Newtonian mechanics have also been rather extensively developed (e.g., in the general theory of relativity).

The Kantian theory of the a priori, then, raises a problem that Kant himself could scarcely have considered important, since for him the a priori laws of geometry and kinematics appeared to have but one form. A general methodology of science must include an

[1] In the pure formal theory, the choice of the value of c is arbitrary, and hence there are really an infinite set of Special Theories, if c is infinite, then the classic theory and the Galilean transformations result. For those acquainted with non-Euclidean geometries, c is analogous to certain constants that appear in homogenous geometries with a constant curvature.

account of the manner in which we can investigate the a priori laws that are required to define an observation. Apparently we cannot avail ourselves of the empiricist's appeal to experience in the matter, for the Kantian doctrine declares that any such appeal to experience, since it is a process of learning by experience, must presuppose a geometry and a kinematics.

Many have been inclined to reject the Kantian theory of knowledge on the ground that it provides us with no way of determining the correct a priori systems of geometry and mechanics, but such a rejection is really impossible. Kant's great contribution to the theory of scientific method consisted in pointing out the necessity for assuming *some* formal system in an experimental investigation: no investigator goes to his experiment without presuppositions. We cannot turn back and forget the lessons history has taught; the Lockian notion of a perfectly open (and blank) mind is an impossible one to conceive if we also demand that such a mind shall learn something from its experiences; according to Kant, an "open" mind could gather at best only the immediate data of the senses, and would have no way of coordinating or unifying these data. No, we cannot reject the Kantian doctrine of the a priori simply because it presents difficulties its author could not have realized; a scientific method must be found that will retain the a priori and also provide a way of testing its validity.

One method for determining a priori truth will occur to the reader immediately; certain of Kant's passages suggest the method, and the idea was developed in the writings of Herbert Spencer. In making the distinction between the a priori and the a posteriori, Kant says: "The question now is as to a *criterion*, by which we may securely distinguish a pure from an empirical cognition. Experience no doubt teaches us that this or that object is constituted in such and such a manner, but not that it could not possibly exist otherwise. Now . . . if we have a proposition which contains the idea of necessity in its very conception, it is a judgment a priori; if, moreover, it is not derived from any other proposition unless one equally involving the idea of necessity, it is absolutely a priori" (2, Intr., Sec. II). The quotation suggests that intellectual "intuition" provides a way of deciding a priori laws. The suggestion was carried out in Spencer's

First Principles. Following Kant (and the empiricists), Spencer asserts that the mind has certain sensations ("manifestations") given it by the "Unknown" external world. The coordination of these takes place by means of the principles of Space, Time, Matter, Motion, and Force. Spencer lists what he considers to be the fundamental principles of the Known: the indestructibility of matter, the continuity of motion, the persistence of force, etc., ultimately based on his famous evolutionary principle. The "critical" attitude is expressed in the following: "It must be added that no experimental verification of the truth, e.g., that Matter is indestructible, is possible without a tacit assumption of it. For all such verification implies weighing, and weighing assumes that the matter forming the weight remains the same" (6, p. 182). Again, "from the standard-measure preserved at Westminster are derived the measures for trigonometrical surveys, for geodesy, the measurement of terrestrial arcs, and the calculations of astronomical distances, dimensions, etc., and therefore for Astronomy at large. Were these units of length, original and derived, irregularly variable, there could be no celestial mechanics nor any of that verification yielded by it of the constancy of the celestial masses and of their energies. Hence, persistence of the space-occupying species of force cannot be proved [by experiment], for the reason that it is tacitly assumed in every experiment or observation by which it is proposed to prove it" (6, p. 197).

The necessity of assuming a certain principle in order to conduct experimental investigations might seem to be proof enough of its validity for the experimenter, were it not that some more ultimate proof of this necessity is required. Why must the experimenter presuppose *this* particular law? Spencer answers: "Our inability to conceive matter becoming nonexistent is consequent on the nature of thought" (6, p. 181). Such passages indicate that the validity of a priori laws rests on intuition for Spencer. The similarity in this respect between the development of rationalism and the development of criticism is noteworthy, for the classical rationalist (e.g., Spinoza) felt obliged to base his beliefs on certain unproved and unprovable statements; similarly, of all the a priori laws Spencer states: "All reasoned-out conclusions whatever must rest on some postulate. We cannot go on merging derivative truths in those

wider truths from which they are derived, without reaching at last a wider truth which can be merged in no other, or derived from no other" (6, p. 199).

The arguments one finds so forcible against the case of the intuitive method as employed by the rationalist seem to be equally forcible here. If the validity of a priori laws is to rest on an intuition or awareness of their truth, then the critical method has become a specialized one, applicable (at best) only by those who feel the intuition. No directions or method can be provided for the employment of this intuition, and he who is at a loss to understand its meaning must give up the hope of discovering truth. The criterion of generality is in opposition to the intuitive method.

Just as the modern rationalists feel obliged to abandon intuition and employ a truth-by-definition method based on pure convention, so some of the later criticists attempted to solve Kant's fundamental problem by arguing that the choice of a priori law was "arbitrary." This solution is presented nowhere so forcibly as in the writings of Henri Poincaré, of which the following passage, on the question of time, is representative.

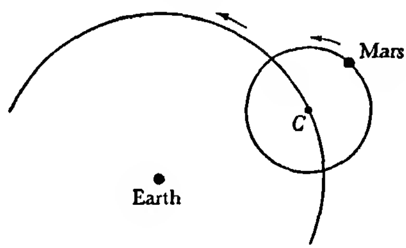
> We do not have a direct intuition of simultaneity, no more a direct intuition of the equality of two durations.
>
> If we believe that we have this intuition, it is an illusion.[1]
>
> We gain the idea of simultaneity by means of certain rules which we employ continuously without explaining why.
>
> But what is the nature of these rules? . . .
>
> These rules are not forced on us and we can amuse ourselves by inventing others; however, we cannot deny the rules without complicating to a large extent the formulation of the laws of physics, of mechanics, of astronomy.
>
> We choose these rules, then, *not because they are true, but because they are the most convenient,* and we could sum them up by saying:
>
> "The simultaneity of two events, or the order of their succession, the equality of two durations, ought to be defined in such a manner that the formulation of the natural laws is as simple as possible. In other words, all these rules, all these definitions, are only the fruit of an unconscious opportunism" (5, Chap. II, Sec. XIII).

A familiar illustration will help in the understanding of Poincaré's position. One might feel inclined to assert that the laws describing

[1] Thus Poincaré denies Spencer's solution.

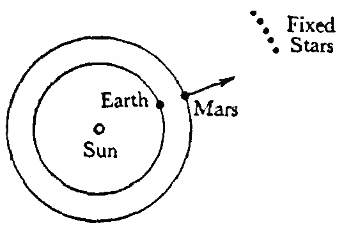
the motions of planets were a posteriori, and determinable only after the observations of the various planetary positions; but our illustration of Kepler's determination of the position of Mars, given above, shows that at least a portion of astronomy must be assumed a priori, and we must inquire what scheme the astronomer is to assume. If we use Poincaré's criterion, that system of planetary motion should be assumed which provides us with the simplest and most convenient formulation of all astronomical laws. This point may be illustrated by a brief reference to the early history of experimental astronomy.

In the Ptolemaic scheme, the earth was taken as the center of the universe, and, specifically, the center of the solar system. The positions of the stars at various times were then determined by assuming this fact and making use of the "fixed stars" as a background. Certain philosophical theories, and perhaps the motive of simplicity, led to the assumption that the planets traveled in circles about the earth; but now the observations made on the motions of the planets caused difficulties. If



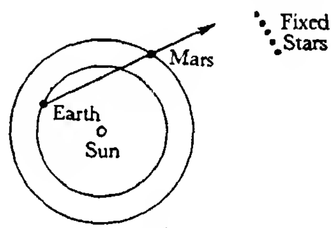
one were watching the behavior of Mars against the backdrop of the fixed stars, for example, the planet appeared to progress during a certain period, then stop, and, amazingly enough, retrogress in the direction from which it had come. This retrogression ultimately stopped and the planet again progressed. The Ptolemaic system explained this phenomenon by means of *epicycles*, picturing the planet as traveling around the circumference of a relatively small circle whose center traveled on a larger circle around the earth (see figure). The use of epicycles did permit an explanation of the planetary retrogression (since the planet was carried backwards for a certain period by its epicycle), but further complications were necessitated by more accurate observations; the center of the epicycle, for example, had to be placed off the circumference of the large orbit, giving an "eccentric" form of motion. The analytical equation of planetary motion under the Ptolemaic scheme would be very complicated.

The "Copernican Revolution" consisted in assuming, as an a priori law, that the sun, and not the earth, was the center of the solar system, and that the earth and the remaining planets traveled in circular orbits. The complications of the Ptolemaic scheme miraculously disappeared. The epicycles were no longer necessary (or at least, not so many of them), for now the retrogressive behavior of the planets was explainable in terms of the motion of the earth (and the observer). For example, since the earth is closer to the sun than Mars and completes its orbit in a little more than half the time required by that planet, there will be a period when Mars appears to travel backward, just as a slow train, actually traveling in the same direction as an observer on a fast train, may appear to



I

Mars appears to an observer on the Earth to be retrogressing at this point



II

A half-year later, Mars has progressed for an observer on the Earth.

him to be traveling in the opposite direction. During another period, the planet will appear to be progressing, because of the position of the earth. The diagrams illustrate these results.

Copernicus provided a very simple solution of the problem of planetary motion, but the relatively exact observations of Tycho Brahe, analyzed by Kepler, indicated that the Copernican system would have to become more complicated to explain certain observations. Kepler's solution in effect still further simplified the astronomical theory by attributing to the planets the elliptical orbits that swept out (from the sun) equal areas in equal times. The Copernican solution really became a special case of the Keplerian, the circle being a special case of the ellipse. But even Kepler's solution was not the most simple and convenient, for, in the first place, it offered no connection between the motions of the planets

and the motions of bodies on the surface of the earth, and, second, it also would have to be complicated in the light of more observations. Newton's solution consisted in removing both of these difficulties in the Keplerian system. He devised a general law (the Law of Universal Gravitation) describing the manner in which all material particles attract one another proportionally to their masses, inversely as the square of their distance apart. Though as a consequence, the planets no longer traveled in the path of a true ellipse because of the attraction of one planet on another, the general equations of motion were very simple, and with the aid of the methods of differential and integral calculus, were mathematically rather easy to handle.

It is true that many empiricists have also used the terms "simplicity" and "convenience" in discussions of the role of law in science. A word of comparison will help to emphasize the basic difference between criticism and empiricism; the difference is one of "presuppositional" priority. According to Pearson and all statistical empiricists, questions about laws of nature can only be answered by presupposing certain facts: the scientist has interest in laws as ways of conveniently *describing* observations previously made. But for Poincaré, no observations can be made without presupposing certain laws; the facts of experience are meaningless unless interpreted in a presupposed space-time framework, and the best presuppositions are those that provide the simplest and most convenient interpretations. Thus, the empiricist pictures Kepler (or Tycho Brahe) as mapping the positions of the planets in the heavens by "direct" observations, and then inquiring what path provides the simplest description of these positions. If this were the true account of Kepler's method, one is astounded that it took so long to discover that the path was elliptical. The criticist, on the other hand, asserts that the figures gathered by Tycho Brahe on the various angular positions of the planets could not even have been gathered if some geometry had not been presupposed, and once gathered, would have been utterly meaningless if some astronomical scheme had not been presupposed. The scientist cannot begin his researches without a formal system of laws, and of the many possible formal systems, he chooses the simplest and most convenient.

Poincaré's solution of the problem of the correct a priori cannot be said to be adequate until the rather indefinite terms "simplicity" and "convenience" have been defined. But before proceeding to the difficulties involved in such defining, we should make a remark concerning an oft-repeated phrase in Poincaré's writing; we choose certain laws, he asserts, "not because they are true, but because they are the most convenient." If this is the case, one is inclined to ask what is the distinction between the *truth* of an assertion and its *convenience* for the scientist. If a priori truth is not determined by criteria of convenience, then by what criteria is it determined? In other words, if there is a distinction between the truth and the convenience of an a priori proposition, then surely the scientist should accept the true a priori, and not, as Poincaré insists, the most convenient. Poincaré prefers to say that there is no a priori truth. But there is a method for selecting the a priori. This seems curious methodology: there are criteria of preference among a priori principles, but no truth. Rather, why not say: "We choose certain laws because they are the most convenient, and *therefore*, because they are true." Convenience and simplicity then become the criteria of a priori truth, and the methods used to determine convenience and simplicity are really methods to determine truth as well.

In order to comment on these criteria of the a priori we may summarize the critical analysis of experimental method by making special reference to the techniques of inference described in Chapter II. These techniques of inference demanded that a question be posed as a set of alternative hypotheses. These are assertions common to all hypotheses, and these we have called the "presuppositions of inquiry"; the remaining assertions, which change from hypothesis to hypothesis, are called the "basis of inquiry." Within the terminology of Chapter II the presuppositions were statistical in their language: they asserted the manner in which the observations were drawn, the universe from which they were drawn, etc. The critical analysis of this chapter is designed to show that these presuppositions must be considerably extended beyond the purely statistical if we are to make observations that are pertinent in the answering of a question. In particular, we must specify certain very general characteristics of the natural universe in which

the observations were made. We must specify the space-time properties of the natural universe, as well as the general laws governing the regularity of the motions of particles. We cannot expect first to answer questions about these general aspects of nature, for, like the statistical presupposition of randomness, such answers must be presupposed in all cases except the sense immediacies; and the sense immediacies are too poor in connotation to provide a basis for the answering of general questions.

The *generalized* analysis of methodology now takes the following form. In order to answer questions about the natural world we must perform the following operations:

1. A theory of space, time, kinematics, mechanics, and probability must be constructed. These theories are formal in the sense developed within modern rationalism. They must be rich enough to specify how an observation is made, i.e., the conditions under which observations are given meaning; in particular, the theories must enable the experimenter to *individuate* and to *identify* objects in the natural world.

The theories in general must be rich enough to specify the conditions under which a set of observations are *pertinent* to the answering of a question. For example, if we ask how heavy a certain object is, we require confirmation that a certain set of operations involving an analytical balance will provide pertinent data that can be used to answer the original question. In other words, the operationalist (empirical) philosophy must be generalized; the meaning of a question, or concept, depends on a set of operations, it is true, but the dependence is not arbitrary. Whether or not a specific set of operations provides pertinent data depends upon what kind of natural world we presuppose.

This is not to say that every experimenter is aware of constructing a natural image sufficient to guarantee the pertinence of observation. But what he is unaware of, but must nevertheless do, he does unconsciously, i.e., intuitively, and in this respect less efficiently in the long run.

One might feel that a geometry, kinematics, and mechanics are *sufficient* to guarantee the pertinence of observation. Our subsequent analysis will show that they are indeed *necessary*, but that a

far more general knowledge than is provided by these three sciences is required for the sufficient criteria of pertinence.

2. A set of pertinent observations are then made, under the conditions specified by the general theories of nature. Observation, for the critical philosopher, involves a dual aspect: the immediate data of sensation and the generalized construct provided by the theories presupposed under 1. That is, observations are not purely theoretical entities; they involve, besides, what is "given" in sensation. Hence, there is at least one aspect of nature which we cannot investigate by the methodology we are outlining. The meaning of sense intuitions, i.e., the answering of questions about such intuitions, does *not* depend upon the general presuppositions about nature required by 1. The meaning of such intuitions is given immediately, and we run no risks in choosing an answer to the type of question they raise. Sense intuitions provide the ground for the distinction between the *theoretical* and the *practical*. Those questions which demand such intuitions for their answering are called practical, and the answers to such questions provide objective knowledge about the natural world.

3. A set of alternative hypotheses must be constructed. The presuppositions of these hypotheses, i.e., the assertions common to them all, are the theoretical laws of geometry, kinematics, mechanics, and probability theory contained in 1. Specifically, the presuppositions show how these general theories are applied to the sense intuitions. The presuppositions are therefore a "schema," in Kant's sense, which shows how the complex concept of "observation" is built up out of sense intuition and theory.

4. The method of selection of an alternative hypothesis is not discussed within the classical philosophy of criticism, since this aspect of methodology was unknown in Kant's time.

The history of philosophy has thus forced us to generalize considerably on the method of inference we first presented. The original presentation demanded simply that observations be made; the critical philosophy has shown how much is involved in this demand, and in this sense represents a revolution in methodology as significant as the Copernican in astronomy. We do not derive the pertinence of observation out of nature; rather, we put such

pertinence into the natural world, and if we did not do so, we would be left with a relatively meaningless set of immediate sense data.

That the critical analysis leaves several important problems still to be answered there can be no doubt. For one thing, we want to know whether questions about the general theories of geometry, kinematics, and mechanics must be answered by a technique different from that given above. For another, are these sciences sufficient to guarantee the pertinence of an observation with respect to a given question?

Poincaré's answer to the first of these problems will carry us on to the next stage "beyond criticism." According to his analysis, the laws of geometry are formal in character, and are best described as "conventions." Thus, Euclidean geometry provides a convenient model for individuation. Now it was apparently Poincaré's belief that the choice of convention was in some sense arbitrary, and lay with the individual scientist. This is a feeling commonly enough expressed in the scientific literature: it is assumed that if a given proposition can be designated as a "convention" of the science, then one has automatically removed all problems concerning its truth. Yet the history of science has shown that the so-called conventions of yesterday become the falsities of today. There is a real sense in which we can say that the Euclidean geometry does *not* adequately describe the physical space of the universe. Put otherwise, as observations are collected, we are forced to revise our opinion as to what is convenient and simple; apparently the a priori laws are no "safer" in this respect than a posteriori. An experiment that is designed to check the validity of a certain theory actually checks (or refutes) the validity of the presuppositions which the experimenter made. This turn of events should not prove very surprising to the formal scientist: he is aware that in general it is a fairly simple matter to interchange the role of *definition* (or convention) and *postulate* in a formal system. The postulates may be made to become definitions, or "truths by convention," by a transformation of the undefined concepts. And hence, if a physicist is subjecting a formal system to the test of experiment, he is submitting both the postulates and the "conventions" to this

test, and the two play very much the same role as far as the final results are concerned.

The consequence is that the words "simple" and "convenient" lead to a reinterpretation of the actual role played by the a priori laws; simplicity and convenience depend for their meaning on other information: one cannot decide without considerable experience in the physical sciences whether a given formulation of a law is a convenient one. Hence, the verification of an a priori law on the basis of convenience must depend upon other information, and the a priori has lost its priority.

The story of criticism then points out an important moral: the scientist should always beware of basing truth on convention, beware, that is, of using a definition to resolve a problem. In this connection, consider the last and still unstudied aspect of the methodology of inference, the selection of one of the alternatives. There are a number of methods for choosing among the alternative hypotheses; the question of fundamental importance is therefore to select the "best" method. What does "best" mean? Is not the choice of definition of best test an "arbitrary" one, or can we say that there is a "true" definition of best? To say that the choice is arbitrary is absurd, for then we give up the hope of selecting a general criterion of precision in experimental method. Yet to say that there is a true definition also seems absurd, for we then require some method of determining whether a given definition is true. And if, as appears reasonable, this method is to depend upon the social consequences of choosing one definition rather than another, then we seem to have involved ourselves in a circle: to determine the social consequences, we must conduct an experiment, make observations, and analyze the results. But the analysis of the results must be made, presumably, by the best method.

Consequently, we now seem forced to the following paradoxical conclusions:

1. We must agree with Kant that the construction of an intelligible world out of the immediacies of sensation demands a certain a priori equipment, i.e., demands the assumption of certain principles not derivable from the elementary facts of experience.

2. But we must disagree with Kant that a priori laws are known intuitively, and that their verification is independent of observation.

Thus, our story seems to have led us to what many would be inclined to call an *impasse;* we are required to assume a priori law to learn about the world, but we cannot determine the correct a priori until we have learned about the world.

The post-Kantian nineteenth century development in the philosophy of science was an effort to find a solution to this *impasse.* The initial phase of the development consisted of a return to the analysis of the *immediate* in science, and this study will carry us to the next school of philosophical thinking on methodology.

REFERENCES

1. Einstein, A., *Relativity*, Methuen (1900).
2. Kant, I., *Critique of Pure Reason* (2nd ed.), translated by J. M. D. Meiklejohn, George Bell (1884).
3. Mill, J. S., *A System of Logic*, 8th ed., Longmans Green (1872).
4. Poincaré, H., *Foundations of Science*, The Science Press (1907).
5. Poincaré, H., *La Valeur de la Science*, Bibliothèque de Philosophie Scientifique, Flammarion (1912).
6. Spencer, H., *First Principles*, Appleton (1900).

Chapter IX Relativism

Our discussion of answers to the problem of truth has shown that the solution depends on deciding the answer to another problem, namely, where does science *begin* in its investigations, or, in our terminology, what are the questions whose answering presupposes no prior knowledge? The rationalist proposes that this starting point must be certain laws, which, since they are known without presupposition, are known *intuitively.* The empiricist, on the other hand, proposes factual statements as the beginning of all knowing. A review of our study into the nature of such "initial facts" will be helpful here.

An analysis of the factual data of science is apt to be slighted in the writings of many empiricists, even though the importance of such an analysis is all too apparent. Indeed, the significance of research conducted by empirical methods is apt to be negligible if a faulty method has been employed in gathering data. For example, everyone would admit the absurdity of having a color-blind person collect data on the distribution of red corpuscles in a liter of blood; we would refuse to grant that the data had any value in the formation of statistical inferences on the subject. Similarly, we might question the statistical inferences concerning the ill effects of smoking based on data collected by a religious group, one of whose moral tenets forbade this indulgence. The suspicion of the average man concerning statistical inferences is probably based in large measure on the lack of assurance that the data are "fairly" collected, i.e., are pertinent to the answering of the questions. When a famous poll of public opinion sends men out to determine the number of people in favor of a certain government policy, may not

the resulting "data" vary depending on the mood of the inquirer, the way in which the words of the question are accented, and so on? Statisticians, conducting a "sampling" of a universe, demand that the cases chosen shall be "random," and admit the inadequacy of the sample if random choice is absent. It becomes clear, then, that careful methods must be outlined for the gathering of data. We cannot require the investigator to be unprejudiced, openminded, unambiguous, unless we define these terms, i.e., provide a method for determining the existence of prejudices, lack of openmindedness, ambiguity. In general we must have a method for deciding whether a set of data are pertinent or not with respect to a certain problem.

The need for a careful analysis of statistical data indicates that the data are really *inferences* from more basic ones. Thus we may *infer* that if we hear someone *say* he disapproves of a government policy, that he actually does disapprove; we may *infer* that if a man *says* he has pains after eating a certain food, he actually does have pains. Such inferences are usually based on more than one fact; we infer that X has a pain because (a) we hear him say so, (b) his face appears contorted, (c) his body appears to have a certain muscular tension. It therefore becomes the task of a general scientific methodology to set down the conditions for *correct* inferences of this sort. It also becomes the task of a general scientific method to make explicit what are the beginnings of all such inferences. The empiricist, then, has been driven to the same problems as the rationalist: he must show the rules for correct inferences, and he must show what are the basic premises from which all such inferences are to proceed; like the rationalist, he refuses to accept the infinity of problems that must arise if we admit that such inferences have no beginning. The method of the two schools is the same; they simply differ on what shall be the beginning point, fact, or law.

A fundamental problem, then, of empiricism, like that of rationalism, is a search for those facts not inferred from anything else: "The logic of observation, then, consists solely in a correct discrimination between that, in the result of observation, which has really been perceived, and that which is an inference from perception." And what remains when all inference is taken away? "There remain, in the first place, the mind's own feelings or states of con-

sciousness, namely, its outward feelings or sensations, and its inward feelings — its thoughts, emotions, and volitions" (12, Bk. IV, Chap. I, pp. 188–189). The uninferred beginnings of all fact-finding are simple sensations and reflections.

But as our discussion moved on, it became apparent that Mill's (or Locke's) rather naive account of these ultimate data was not adequate. Not *all* my sensations are ultimate, unanalyzable, immediate; when I have a sensation that a book is on the table, I compound (or infer) one piece of sense-data from other more ultimate ones, as Kant (criticism) was so careful to point out. Criticism reduced the immediate data of sensation and reflection to things much more simple than the earlier empiricists had thought them, to things so simple, indeed, that no inferences could be made from the ultimate data without the aid of laws.

Indeed, the analysis of criticism of the immediate data of sensation implied that such data were meaningless for the scientist, though the proponents of criticism seem in general to have missed the implication, for it was shown that *no* proposition containing a relation could be immediately known. But the statements "I see red," "I feel a pain," are propositions containing relations — relations holding between an observer and his observation. And we cannot even say "there is a sensation of pain" unless we specify the conditions under which the form "there is an X" holds. But such *general* specifications determining the truth of a statement make that determination nonimmediate, for in order to answer the question whether the given statement is true I must first of all determine whether it accords with the general specifications.

In order to make the argument against the meaningfulness of immediate sense-data clearer, we may turn to the lessons history has taught us of the cost of maintaining such data.

1. The sophist Gorgias is supposedly responsible for the famous nihilistic trio of statements: "Nothing is," "Even if something were, it could not be known," and "Even if something could be known, it could not be imparted to others." It is the last of these tenets we wish to discuss. If we reconstruct Gorgias's argument on the grounds of empiricism, it would run something as follows: If it were possible for me to know something, it would have to be known in

terms of my immediate sense-experience; but in order for me to impart my knowledge to you, I cannot use these immediate data of sensation, and must use some language-symbol to represent them; thus, in order to tell you that the moon is yellow, I cannot transfer my (immediate) sensation of yellow to your mind, but must use (say) English terms to represent the sensation; but then what evidence do I have that these terms will call up in your mind the very image I observed? Perhaps, for all that our methods of knowledge can tell us, you picture things in an entirely different way than I do, so that what appears to me to be yellow, appears to you to be (what I would call) red. And we cannot appeal to the consistency of behavior on our parts in the face of the same stimuli; for the same behavior (at best) would simply mean that there is a consistency in the sequence of sensations in both of us, not that the sensations were the same. But in truth what I call *your* behavior is again a part of my sensations, and is no evidence as to how your behavior looks to your mind. Even our criteria of regularity of behavior may be entirely different.

2. But the inability of one mind to transfer its thoughts to another is not the only consequence of a theory of the immediacy of sense data; there is also the inability of one mind to prove the existence of another. Nay, not only can we not *prove* another's existence, we cannot even show (on the grounds of empiricism) that the hypothesis that other minds exist is even more probable than not. Some empiricists, indeed, have attempted to show that a type of probability judgment can be established by an analogy argument, which runs thus:

> I am aware, and I alone am aware, that certain of my bodily acts are accompanied by mental states. When I observe similar acts in other bodies I infer that they too are accompanied by like states of mind. No experience can be brought to confirm this inference, but then nothing can transpire to refute it (15, p. 3).

But the fallacy of the analogy argument is readily shown:

> The analogy argument calls its procedure an inference. Now, everybody knows an inference from a thousand cases to be more valuable than one drawn from a hundred, an anticipation based on a hundred observations to be safer than one with only ten to support it. But there are

these who, knowing all this, would conclude that an inference from one instance has some value. If in my case mental states accompany my body's behavior, there is at least *some* ground for supposing like acts of another's body to be in like manner paralleled.

This illusion, for it is one, springs I think from a failure to catch the meaning of inference. An inference from a single case, if it be really an inference from a single case, has exactly no value at all. No one would be tempted to attribute eight planets to every sun because our sun has eight such satellites. The reason a single observation is sometimes correctly assumed to have weight is that the method of observing has been previously tested in a variety of cases. The shopkeeper measures his bit of fabric but once; he has, however, measured other fabrics by the same method numberless times, and has a fairly clear idea of the probable error of his result. But the principle holds absolutely of all results: no series of observations, no probable error; no ground for inference; no meaning as a datum[1] (15, pp. 3-5).

Thus another cost for the empiricist in maintaining immediate sense-data is his inability to disprove solipsism.

3. Similarly, the entire argument of Berkeley concerning the inability of the inquiring mind to prove the existence of an "outside" material reality depends on the presupposition that the only immediately knowable propositions are ones concerning our sensations. All (empirical) "idealisms" and "egocentric predicaments" presuppose that such statements as "I am having a certain feeling" or "I am thinking a certain thought" are back of every scientific method, and are the certainties of such methods: "I know that I have ideas and sensations, but what causes them, what a world without them would be like, I cannot say" or "The only immediately knowable reality is my own mind."

4. Finally, we may raise two additional questions: how does the observer determine the existence of his own mind, and (finally) how does he determine the existence of his own perceptions? In reply to

[1] It might be noted in passing that there is an attempt to reconstruct the analogy argument by asserting that if X and Y have many properties in common, then it is highly likely that they have all their properties in common. But this fallacy is as bad as that already described Are properties random variates? If so, what is their distribution? Does this distribution include the property of consciousness? If so, on what grounds? If the grounds are not given, then there is no case. If the grounds are presupposed, the argument is circular. More practically, those proposing this argument should read into population sampling, where it has been shown over and over that an inference from observed properties of one sample, to another sample having only some of the properties, leads to unwarranted results

these questions, one might feel inclined to use the Cartesian argument, at least to the extent of proving that one thinks. But this argument contains a rather subtle difficulty; I am urged to attempt to doubt the statement "I think," with this consequence: "If I doubt that I think, then I must be thinking, since doubting is an act of thinking." Hence (granting the universal that doubting is an act of thinking, which is possibly a truth of definition), the statement "I doubt" implies "I think." Hence "I think" can be proved if we can prove "I doubt." Now, runs the argument, "I doubt" must be true, because it is either doubtful or not (to me). If it is doubtful to me, i.e., if I doubt its truth, then I must be doubting and hence I reaffirm its truth; if it is not doubtful, it is true, and hence again its truth is reaffirmed. But a closer analysis leads to a third alternative: perhaps the statement "I doubt" is neither doubtful nor certain to me. This third alternative is possible if I am not a thinking being at all (e.g., to a stone a statement is neither doubtful nor certain). But the argument being designed to show that I am a thinking being, we cannot reject the third alternative without begging the question or using another argument.[1]

In the end, all proofs of the existence of an ego, or its perceptions, must rest on statements not requiring further proof, i.e., statements "intuitively" known. The empiricist does not advance beyond the rationalist with respect to the beginnings of science; for both, the elementary propositions of science are known without any method being applied to determine their truth. It seems to matter little whether we say: "Intuit the general principle that 5+7=12," or "Intuit that you are now seeing red." And just as the rationalist explains all words in terms of concepts intuitively known, so the empiricist, in attempting to define his terms by means of "physical operations," reduces them to intuitively known ideas. For the empiricist, everything should be defined by or reducible to immediate sensations, but what the immediate sensations are he cannot say. Empiricism contains all the "mystical" elements of rationalism,

[1] In the logician's terms, the *petitio principii* of Descartes' reasoning lies in the assumption that the "I" must belong to the universe of discourse in which "doubtful" and "certain" are contradictory terms, i e , to the universe of discourse of reflective minds; the assumption is actually equivalent to the thing proved. It might be mentioned by way of historical clarification, that it is doubtful whether Descartes himself thought any logical analysis of his argument was necessary

and he who is not so gifted as to know in every or even some cases what he feels and senses must be forever excluded from the select realm of scientists. Empiricism fails to give a criterion of generality in just the same manner as rationalism fails.

Better to phrase our objections to the immediate data of sensation, we turn to passages of those who have made these objections a focus-point of their thinking:

1. There was at least one careful student of Kant's writings, Solomon Maimon, who felt that Kant had left in a problematical state the important problem of the "given" in experience. Maimon asked the very natural question concerning the nature of the "given," and came to the conclusion that the search for the first given in sensation must be an endless one; we can by analysis restrict the meaning of these "givens," but at any finite stage this analysis is incomplete, for all such "meanings" depend on the categories of the understanding and hence are not primitive.

2. Hegel's critique of Kantianism took the form of showing that the reflective mind can never be satisfied with anything fixed and immutable in the process of learning. This viewpoint is developed (8, pp. 90–103) by showing that every proposed "fixity" (e.g., the fixity of immediate sense data, the fixity of an aesthetic intuition, etc.) leads to its own negation and forces one to seek for truth in other and opposing quarters. From our point of view here, the critique of the immediately given (*sinnliche Gewissheit*) is pertinent. If we do not pervert his meaning, Hegel shows that every immediate judgment of sensation involves (among other things) a "now" and an "I" (*Jetzt und Ich*); e.g., a judgment may be "*I* am *now* seeing red." Unless these two terms are present in such judgments, they cannot be made. But an examination of the language of all such judgments shows that these same terms appear in them all ("I am now hearing a noise," "I am now feeling a pain," etc.); consequently, the terms are not particular, as are the terms representing the sensations, but are *general*, i.e., applicable to an indefinite number of instances; then, like all general concepts. these terms demand rules for their application; but this is exactly what the designation of "immediate judgment" forbids — we must decide without the aid of presupposed rules that "I am now having a sensation X" is

true. Rephrased, the argument would go something as follows: To make the judgment "I am now seeing X," I must decide among other things that the sensation has actually occurred, and such occurrence implies a time-individuation ("at the point 'now'"); but such time-individuations imply a general rule that by the nature of the case cannot be applied. A similar argument is constructible concerning the "I" or "Ego"; the identity of the "I" in all such immediate judgments implies that the ego is not peculiar to the given perception, but is universal. But then we must have a method of deciding whether the given sensation is to be placed in the general class of all my perceptions. Hence follows the well-known Hegelian principle "Simple immediacy is itself a concept of reflection."

The difficulties of Hegel's argument from our point of view are manifest: his inferences are difficult to follow and because of his terminology it is not easy to determine what the bases of these inferences are. Further, one might deny the necessity of a "now" and an "I" in every immediate judgment of sensation, and regard the immediately given as a *quale* — a "something-I-know-not-what." This was certainly Kant's way of regarding the immediately given; the empirical "intuitions" where originally independent of both the intuition of time and the synthesizing function of the Ego; certain modern logicians might phrase the matter by asserting that the immediate in sensation cannot be a "proposition" in the logical sense.

The clearest formal arguments opposing immediacy as a part of scientific method are to be found in Singer's writings; and most representative, perhaps, is the argument to show that the "explanation" of sense-immediacies in terms of qualities is inadequate:

> Modern psychologists generally have not failed to see that a difficulty is contained in any view of sensation which regards it as an immediate datum. They are, for the most part, agreed that sensation so viewed could no longer mean a sound, or a color, or anything else it has meant or may now mean in the language of common sense. A certain compromise is frequently attempted, which consists in stripping sensation of its so-called properties, thus forcing it to stand for "mere quality." Helmholtz, for instance, proposes, in the cause of immediacy, the following criterion of sensation: "No sensation indubitably present could be set aside and destroyed by an act of the understanding; hence, nothing in our

sense perceptions is to be recognized as a sensation which, by momenta evidently derived from past experience can be corrected, in perception and changed into its opposite" (9, pp. 610 f.). In other words, a sensation pure and simple must involve no judgment based on past experience and liable to error, but must be immediately known and recognized. One is not surprised when one finds the application of this criterion driving Helmholtz to the conclusion that "only the qualities of sensation are to be regarded as really pure sensation." . . .

Criticism must take the same form in treating of quality, as it did in dealing with the supposedly richer term, sensation. Quality, too, if it is to stand for a mere datum, loses first its ordinary meaning, then all possible meaning; whereas if it is to retain meaning, it fails to fulfil the conditions of a merely present experience. For, like the more complex term sensation, the term quality of sensation stands in ordinary use for a psycho-physical and psycho-physiological conception. Our obvious motive for calling certain differences of sensation "qualitative," is to distinguish them from local and intensive differences. So that, ordinarily at least, quality is not a genus of which intensity and local sign are species, but all three are rather coordinate species. As such, their differentiae are usually stated in psycho-physical terms; but even where this is not so, they must be mutually dependent for their meaning.

Still, I confess, it is not uncommon to make quality a more generic term than the others mentioned. Thus it is frequently claimed that all mental differences are qualitative, all physical differences quantitative. I do not know what truth there may be in this proposition, regarded as a statement of fact, after one has defined qualitative and quantitative differences. As it stands, it has the air of offering itself as a definition of these terms. If so, it is evident that one could not hope to distinguish between mental and physical differences until one was able to attach meaning to the physical and mental worlds, a meaning which, I fancy, can lay little claim to immediacy. A definition of qualitative difference in these terms would make quality a highly reflective product, not at all an immediate datum.

If, as a last alternative, one drop the adjective "mental," and define qualitative difference as mere difference; even then it is the experience of *differents* which gives rise to the concept of quality. A *quality* could not then be immediate, and if one should claim that *qualities*, including all their differences, might be so, it is difficult to see how the manifoldness of such an experience could be realized without comparison with other experiences which it in part resembled and from which it in part differed. To say, as Bradley does, that qualities and relations can, under proper conditions, be present, but not recognized, is in violation of his own most cherished principles. It is exactly that divorcement of existence and

recognition, an elevation of "mere possibility" into a place of dignity among meaningful concepts, which he generally condemns with such force and skill in the theories of others.

But further dwelling on this point would not be helpful. Suffice it to say that in proportion as one abstracts from judgments and inferences referring to experiences, possible, but not actual, one abstracts from meaning. So that if one has carried the method far enough to retain nothing that is not immediately given, one must have succeeded in getting rid of all meaning — which, of course, is not the conclusion we wished for. The method must in the end make the immediate quite inarticulate (15, pp. 168-172).

Of course, the reduction of sensations to qualities by no means marks the end of the psychological story. The general import of modern experimental psychology, together with its theoretical background, seems to be that sensation, far from being a simple element of knowledge, is as complex and difficult to understand as are the physical elements (1, pp. 17–22). The dispute between the associationists and gestaltists has led to a major revision of the Lockian analysis. For gestalt psychology, the individual does not observe the simple elements and construct the whole pattern of his sensory field from these; rather, the "simple" element is the result of an analysis on the individual's part, and far from being simple is actually one of the most refined and complex of his ideas. Though the associationists do not go along with this description, they have been forced by the gestalt experiments to regard the elements of sensation as very complex entities. And along with all this, there has been the less systematic psychoanalytic theory (7); here the content of consciousness is only a small part of the individual's mind, and by no means the most significant part. Further, what comes to consciousness is by no means "simple"; it is a result of complex forces representing instinctual desires, social compulsives. Also, what an individual will let himself realize about the content of his conscious mind is again a result of these forces. Hence, under psychoanalysis, the introspective analysis of one's mind, far from being an accurate picture of the content of consciousness, is rather data for the analyst's inference to repressions, inhibitions, etc. Thus, the early attempt to show the ultimate simplicity

of the "given" in sensation has ended within this era in a confusion of scientific theories, from which one can gather at least one result: the "simple" sensation of Locke has gone the way of the "simple" atom of Democritus. And with the passing of the simple sensation, one hopes, will go the "simple intuitions" of the introspective philosophers. I have reference to the "it-seems-to-me" type of analysis of certain British and American thinkers. It is the type of philosophy that bases its conclusions on the philosopher's individual failure to see how such-and-such can possibly be so. Such writings may make good material for the psychoanalyst, but why publish so much of it?

It is true that one may attempt to escape the awkward features of the immediacies that we have been emphasizing by refusing to regard an immediate sensation as a propositional form. In this case, the objection that simple immediacy must be inarticulate has no force, since the sensation does not express any assertion; it is merely a "thing," or, more exactly, it belongs to a different universe from the universe of propositions. But such an escape also ignores the problem at hand, which is to decide what factual propositions form the beginning points of our knowledge; so that if immediate sensation does not form such a beginning point, then we have a right to inquire what does. If the beginning point is a process of articulation of a certain sensation, then it is this process which is the object of our examination, and all the arguments we have gathered against immediacy must be conclusive in showing that such a process cannot be an absolute beginning of science.

But it would be a mistake to suppose that in giving up the immediacies we have simplified our problem. Indeed, why should so many thinkers have clung to the awkward concept of an immediately given, if one could easily construct a scientific method without it?

In particular, the attitude that science *does* begin with the data of experience has been based on the desire to make certain important distinctions within methodology. For example, if there is no immediately given in observation, then how shall we distinguish between the actual and the imaginary within science? We have said that all answers to questions demand a set of observations; if we remove the necessary conditions that observation be grounded in immediate

sense experience, then how shall we guarantee anything empirical in our knowledge? Will not experimental science turn out to be as arbitrary a game as chess? We have talked of the *cost* of maintaining the immediacies; the *profit* seems to be that they enable us to distinguish between the *objective* and *subjective*, the *empirical* and *nonempirical*, the *practical* and *theoretical*. Again, without the aid of the given in sensation how could we distinguish between what *actually* happens (in direct experience), and what *ought* to happen (as judged by our ideals)?

The philosophical viewpoint we shall call relativism has attempted to show how all these distinctions of classical empiricism can be maintained, even though there are no immediate data of sensation.

The viewpoint of the relativism of scientific method is one aspect of the school of philosophy called "pragmatism." The pragmatic doctrine had its roots in a good part of the American tradition, its chief source being C. S. Peirce (13), but it received its most forceful exposition at the hands of William James.[1] In one of his earlier essays, "The Will to Believe," James attempts to analyze the nature of belief, and comes to the implied conclusion that all beliefs receive their meaning in the actions of the believer; to know what A believes about a certain question, we must note how A acts when circumstances demanding an answer to the question arise. Thus, if I carefully avoid passing under ladders, I must believe in a certain superstition, no matter how vehemently I may deny it. Indeed, a man may be convinced he believes one thing but actually believes another; the lover may argue his indifference to his beloved, but his actions tell us otherwise; an old soldier may urge his dislike of fighting, but his stories and the eagerness in his eyes belie him. The naive assumption that a man may think one thing and do another is untenable for the psychologist. And why? Simply because purely "private" thoughts are never objects of observation, and hence must be meaningless to the scientist. Even the subject himself is never sure how he feels, and the essence of "growing up"

[1] Throughout this discussion, there is no attempt made to identify the philosophers as relativists; the pragmatists, Peirce, James, Dewey, Schiller, and others, seem to oscillate between various schools in their writings Rather, certain passages have been chosen as illustrative of the relativist viewpoint, though not necessarily of the pragmatic

scientifically is to relinquish the idea that you know more about yourself than others do or can. One's own thoughts and beliefs are as "public" as the distance of Mars from the Sun at a certain time; both are "testable" by observing the behavior of bodies. Thus the pragmatic doctrine seems to imply the denial of any immediate sense data, for the feeling I have at the moment can only be determined when my behavior is observed. Further, pragmatism suggests a general scientific methodology; if I am faced with two or more alternative answers to a problem, which shall I choose? My refusal to accept any of the alternatives, in that this refusal is a type of action, will itself be a decision, so that all meaningful problems demand of the questioner an answer, i.e., a mode of action. The term "meaningful" here is taken in the pragmatic sense: a problem is meaningful if its answering affects the behavior of him who asks it. For example, the question whether there exists a world independent of the observer would be meaningful to a thinker only if his behavior would change in answering the question one way or the other. If he answered "no" he might feel inclined to commit suicide, in which case the problem is meaningful, but if his behavior (as is usually the case) remained unchanged, the question would be without meaning for him. Evidently, pragmatic meaningfulness for James is relative to the particular thinker asking the questions.

The questioner evidently has no choice but to assume one alternative as valid, to act as though it were valid, and to continue to act in such a manner until future events indicate a change in belief:

> Now truth is always a go-between, a smoother-over of transitions. It marries old opinion to new fact so as ever to show a minimum of jolt, a maximum of continuity. We hold a theory true just in proportion to its success in solving this "problem of maxima and minima." But success in solving this problem is eminently a matter of approximation. We say this theory solves it on the whole more satisfactorily to ourselves, and individuals will emphasize their points of satisfaction differently. To a certain degree, therefore, everything here is plastic.
>
> The point I now urge you to observe particularly is the part played by the older truths. Failure to take account of it is the source of much of the unjust criticism levelled against pragmatism. Their influence is absolutely controlling. Loyalty to them is the first principle — in most cases it is the only principle; for by far the most usual way of handling phe-

nomena so novel that they would make for a serious rearrangement of our preconception is to ignore them altogether, or to abuse those who bear witness for them.

You doubtless wish examples of this process of truth's growth, and the only trouble is their superabundance. The simplest case of new truth is of course the mere numerical addition of new kinds of fact, or of a new single fact of an old kind, to our experience — an addition that involves no alteration in the old beliefs. Day follows day, and its contents are simply added. The new contents themselves are not true, they simply come and are. Truth is what we say about them, and when we say that they have come, truth is satisfied by the plain additive formula.

But often the day's contents oblige a rearrangement. If I should now utter piercing shrieks and act like a maniac on this platform, it would make many of you revise your ideas as to the probable worth of my philosophy. "Radium" came the other day as part of the day's content, and seemed for a moment to contradict our ideas of the whole order of nature, that order having come to be identified with what is called the conservation of energy. The mere sight of radium paying heat away indefinitely out of its own pocket seemed to violate that conservation. What to think? If the radiations from it were nothing but an escape of unsuspected "potential" energy, pre-existent inside of the atoms, the principle of conservation would be saved. The discovery of "helium" as the radiation's outcome, opened a way to this belief. So Ramsay's view is generally held to be true, because, although it extends our old ideas of energy, it causes a minimum of alteration in their nature (10, pp. 61–63).

Pragmatism developed in many ways, but one of its most important branches was the "instrumentalism" of Dewey. The term is evident enough, for under the pragmatic doctrine, ideas and beliefs are the "instruments" for further action. The relativism of instrumentalism is rather clearly expounded in Dewey's *The Quest for Certainty*. Dewey admits with Kant that "experimentation proceeds on the basis of a directive idea," but argues that "an idea is tentative, conditional, not fixed and rigorously determinate" (6, p. 288). It is true that "traditional empiricism has misread the significance of conceptions or general ideas; it has connected them with experience of the actual world; it has connected the origin — and validity of general ideas with antecedent experience. According to it, concepts are formed by comparing particular objects, already perceived, with one another, and then eliminating the elements in which they disagree and retaining that which they have in common.

Concepts are thus simply memoranda of identical features in objects already perceived" (6, p. 166). Traditional empiricism has thus failed to realize the important role of generalizations; its "ideas are dead, incapable of performing a regulative office in new situations" (6, p. 166). The truly significant role of generalizations and universal ideas is that of a tool or instrument; and Kant was correct in insisting that without the employment of such tools of the understanding, experiences could not be made meaningful. Hence, the gathering of facts does presuppose (as an instrument for the gathering) certain prior knowledge: the answering of any question of fact presupposes the answering of at least some questions of law. But the Kantian error lay in assuming "that this previous knowledge need be immediate or intuitive." Rather, "Objects of previous knowledge supply working hypotheses for new situations; they are the source of suggestion of new operations; they direct inquiry" (6, pp. 186–187). Hence, *what* tools we use to further investigations depends on what our observations in the past have been like: the answering of any question of law presupposes the answering of at least some question of fact.

In sum, "the basic error of traditional theories of knowledge resides in the isolation and fixation of some phase of the whole process of inquiry in resolving problematic situations. Sometimes sense-data are taken; sometimes, conceptions; sometimes, objects previously known. An episode in a series of operational acts is fastened upon, and then in its isolation and consequent fragmentary character is made the foundation of a theory of knowledge" (6, p. 188).

Once one has denied any unconditional beginning to the researches of science, then one feels obliged to ask the pragmatist "Whither now?" Where is science going, what are its goals? If we deny that science even has a firm grip on anything, then are we not asserting that science answers questions only in a relative sense? We are supposed to assume for the time being that such-and-such is so, in order to determine whether other things are true, such determination always being relative. In the pragmatism of James, such determination is relative to the particular observer, his purposes, his environment, his presuppositions. In the pragmatism of Dewey,

it is relative to the society and its purposes. The errors of older ways of thinking were errors based on a faith in the absolute, a faith that is substantiated by no scientific methodology. The terminological difficulties in pragmatism's definition of truth make the best course one of direct quotation. Says James:

Our pragmatist view is that the truth-relation is a definitely experienceable relation, and therefore describable as well as namable; that it is not unique in kind, and neither invariable nor universal. The relation to its object that makes an idea true in any given instance, is, we say, embodied in intermediate details of reality which lead towards the object, which vary in every instance, and which in every instance can be concretely traced. The chain of workings which an opinion sets up is the opinion's truth, falsehood, or irrelevancy, as the case may be. Every idea that a man has works some consequences in him, in the shape either of actions or of other ideas. Through these consequences, the man's relations to surrounding realities are modified. He is carried nearer to some of them and farther from others, and gets now the feeling that the idea has worked satisfactorily, now that it has not. The idea has put him into touch with something that fulfils its intent, or it has not. . . .

This something is the man's object, primarily. Since the only realities we can talk about are such objects-believed-in, the pragmatist, whenever he says "reality," means in the first instance what may count for the man himself as a reality, what he believes at the moment to be such. Sometimes the reality is a concrete sensible presence. The idea, for example, may be that a certain door opens into a room where a glass of beer may be bought. If opening the door leads to the actual sight and taste of the beer, the man calls the idea true. Or his idea may be that of an abstract relation, say of that between the sides and the hypotenuse of a triangle, such a relation being, of course, a reality quite as much as a glass of beer is. If the thought of such a relation leads him to draw auxiliary lines and to compare the figures they make, he may at last, perceiving one equality after another, see the relation thought of, by a vision quite as particular and direct as was the taste of the beer. If he does so, he calls that idea, also, true. His idea has, in such case, brought him into closer touch with a reality felt at the moment to verify just that idea. Each reality verifies and validates its own idea exclusively; and in each case the verification consists in the satisfactorily-ending consequences, mental or physical, which the idea was able to set up. These "workings" differ in every single instance, they never transcend experience, they consist of particulars, mental or sensible, and they admit of concrete description in every individual case. Pragmatists are unable to see what you can possibly mean by calling an idea true, unless you mean that between it as a ter-

minus *a quo* in someone's mind and some particular reality as a terminus *ad quem*, such concrete workings do or may intervene. Their direction constitutes the idea's reference to that reality, their satisfactoriness constitutes its adaption thereto, and the two things together constitute the "truth" of the idea for its possessor. Without such intermediating portions of concretely real experience the pragmatist sees no materials out of which the adaptive relation called truth can be built up (11, pp. 235–238).

Says Dewey:

If we see that knowing is not the act of an outside spectator but of a participator inside the natural and social scene, then the true object of knowledge resides in the consequences of directed action. When we take this point of view, if only by way of a hypothesis, the perplexities and difficulties of which we have been speaking vanish. For on this basis, there will be as many kinds of known objects as there are kinds of effectively conducted operations of inquiry which result in the consequences intended.

The result of one operation will be as good and true an object of knowledge as is any other, provided it is good at all: provided, that is, it satisfies the condition which induced the inquiry. For if consequences are the object of knowing, then an archetypal antecedent reality is not a model to which the conclusions of inquiry must conform. One might even go as far as to say that there are as many kinds of valid knowledge as there are conclusions wherein distinctive operations have been employed to solve the problems set by antecedently experienced situations (6, pp. 196–197).

Any philosophy that in its quest for certainty ignores the reality of the uncertain in the ongoing processes of nature denies the conditions out of which it arises. The attempt to include all that is doubtful within the fixed grasp of that which is theoretically certain is committed to insincerity and evasion, and in consequence will have the stigmata of internal contradiction. Every such philosophy is marked at some point by a division of its subject-matter into the truly real and the merely apparent, a subject and an object, a physical and a mental, an ideal and an actual, that have nothing to do with one another, save in some mode which is so mysterious as to create an insoluble problem (6, p. 244).

In the light of such passages as the above, one might justly question whether James is right in describing pragmatism as: "The only articulate attempt in the field to say positively what truth actually *consists of*" (11, p. 234). Is it unpragmatic to ask what James means by an "experienceable relation," to ask how one is to determine

whether an "idea has put someone in touch with something that fulfils its intent"? Are we nonpragmatic to ask Dewey what the conditions are that enable one to investigate the "consequences of a directed act," that enable one to decide that an operation "satisfies the conditions which induced an inquiry"? Do we go outside or beyond the realm of pragmatic method entirely if we ask how one is to answer questions about the "resolution of a problematic situation"? If so, if these problems are not pragmatic ones, then who would assert that pragmatism says positively what truth actually consists of? No, pragmatism must take the course of explaining, and explaining, mind you, by the pragmatic method of explanation, the terms it employs to describe a scientific method. But here a difficulty arises; the pragmatist has abandoned all "absolutes," all "first-givens," all "indefinables" in his description of the search for truth, so that his basic methods prevent him from making the terms any more explicit than the individual and his environment permit. Pragmatic meaning, like pragmatic truth, is relative to the observer and his social or natural environment, and the methods of answering questions and the resulting answers are correspondingly relative: no questions of fact or law are (absolutely) answerable.

That relativism proposes no easy philosophy can be seen by examining the relativistic reformulation of the methods of Chapter II. It will turn out that it is a very complicated problem to confirm a theory pragmatically. To assert that the "true object of knowledge resides in the consequences of directed action" is a facile way of describing a methodology the detailed application of which we can only begin to formulate.

In relativistic terms, he who wishes to determine whether a certain hypothesis "works out" must construct a theory of nature adequate to account for an individual *and his purposes*. The Kantian a priori had two weaknesses on relativistic grounds; it assumed that the theories we presuppose are fixed, and it assumed that the fundamental principles of geometry, kinematics, and mechanics are sufficient to account for the meaning of observations. Relativism, on the other hand, insists that the laws of the presupposed sciences are not fixed for every observer, and further that we must add to the fundamental categories of knowledge the science of purpose (teleol-

ogy). Only when we view the world as an aspect of our purposive behavior do observations and conclusions have any meaning. If we regard the meaning of observation to be independent of purpose, then we have committed the basic error of traditional empiricism: "the isolation and fixation of some phase of the whole process of inquiry." In order to make this additional category of scientific method a meaningful one, we must specify the kind of teleology that we must bring to our experiments. This specification is especially important in view of the metaphysical connotations of the term "purpose" or "end." As Dewey says, "It is dangerous to venture at all upon the use of the word 'ends' in connection with existential processes. Apologetic and theological controversies cluster about it and affect its signification. Barring this connotation, the word has an almost inexpugnable honorific flavor, so that to assert that nature is characterized by ends, seems like engaging in an eulogistic rather than an empirical account of nature" (4, p. 97). The empirical approach to purpose depends, for the pragmatist, upon showing the intimate connection between "means" and "ends." "All ends are ends-in-view; they are no longer ideal as characters of Being · · ·, but are the objects of conscious intent. When achieved in existence they are ends because they are then conclusions attained through antecedent endeavor, just as a post is not a goal in itself, but becomes a goal in relation to a runner and his race" (4. p. 112). "The final source of the trouble is · · · that moral and spiritual 'leaders' have propagated the notion that ideal ends may be cultivated in isolation from 'material' means, as if means and material were not synonymous" (6, p. 281). Now if the relation of means and end is to be based upon an empirical analysis, we must make explicit exactly how the relationship occurs in the natural world. A hint is provided in the following passage: "A machine turns out a succession of steel spheres, like ball bearings. These closely resemble one another, because they are products of like process. But there is no absolute exactitude among them. Each process is individual and not exactly identical with others. But the *function* for which the machine is designed does not alter with these changes; an operation, being a relation, is not a process. An operation determines any number of processes and products all differing from one another; but *being* a

telephone or a cutting tool is a self-identical universal, irrespective of the multiplicity of specific objects which manifest the function" (6, pp. 162–163).

The empirical relation of means and end then depends upon finding a common function within the diversity of behavior; the end is the common characteristic among such diversity, and end and means are simply correlative terms, meaningful within any experimental methodology. The preliminary task of methodology would, therefore, seem to be one of making explicit within the natural world what the purposes are which induced an inquiry; such a process would demand an examination of behavior in order to determine what was common to all the actions associated with the inquiry. In other words, before we can begin to answer a question, we must construct a natural image sufficient to account for our purposes in raising the question. We must also be able to state what constitutes a pertinent observation from the point of view of our ends-in-view. If our purpose is merely to count the number of voting persons whose names appear in telephone directories, then we can collect adequate information by a random sampling of such directories; but if our purpose is to determine which candidate will be elected, the random samples drawn from the directories are inadequate. Consequently, the ends-in-view must be made explicit enough to determine the operations for collecting pertinent data. On this basis, we can readily make all the distinctions that motivated traditional empiricism to keep the immediacies of sensations. The objective and the subjective, for example, are simply two aspects of the world as we view it teleologically. That is, a subjective question, involving a purely subjective answer, simply lacks a long-run control; the answer at best serves a *relatively* immediate end. Thus, relativism is able to reflect on the meaning of sensation, and to regard observations themselves as aspects of scientific inquiry. There is something in the meaning of an observation which is *relatively* immediate; it is that aspect of observation which satisfies relatively specialized and immediate goals.

The basic category of inquiry for the relativist, therefore, turns out to be the means-end relationship.[1] Inquiry only takes on mean-

[1] A means-end schema for psychology and the social sciences is given in (3).

ing when we view the natural world through such spectacles. Actually, the type of presupposition we make of a mathematical or mechanical sort will depend upon the manner in which we characterize our purposes. So the manner in which we follow out the procedures of Chapter II will depend upon a teleological construct of the world, as follows:

1. Select those principles of geometry, kinematics, mechanics, probability theory, etc., which will best serve the basic ends-in-view which induced the inquiry, i.e., will act as the best instruments for experimentation.

2. Set up physical operations for collecting data which will serve as the best instruments for the goals of the inquiry.

3. The alternative hypotheses are to be those that we regard as *important;* that is, *each alternative should be such that if we adopt it, then some significant difference will occur in our behavior.* Hence, if the acceptance of either of the statements "X weighs 16,000 lbs." and "X weighs 16,001 lbs." will produce the same behavior, then these should not be regarded as *alternatives;* the alternatives should exhaust the (logical) possibilities *that make a difference.* Thus in industrial acceptance procedures, we may define a "good" lot to be one having .01 defectives or less, and a "bad" lot to be one having .05 defectives or more. The acceptance procedure considers only the two hypotheses: the lot is good, or else the lot is bad. From the point of view of the purposes of the acceptance plan, the assertion that the lot contains .03 defectives, say, is "meaningless," in that we are indifferent about the distinction between such a lot and a good lot, or the distinction between such a lot and a bad lot; our indifference is measured by the fact that we are willing to take the same action with a .03 defective lot as we do with a .01 defective, or with a .05 defective: we are willing to discard the lot or accept it. We do not *believe* a .03 lot to be any better than a .01 lot, or any worse than a .05 lot. If we *actually* think that a .03 lot is better than a .05, in the sense that our purposes are less apt to be accomplished efficiently when we use a .05 lot, then we ought to narrow the distinction between the good and the bad. We "ought to" do so, because otherwise the pragmatic prescription of method has been violated.

4. It is difficult to find in the writings of relativists any very useful hints as to the method of selection one should employ in reaching a decision on a question. The pragmatists are apt to regard the adequate solution of a problem to depend upon whether or not "the trouble or doubt which evoked inquiry will be resolved." But such resolution can only take place with a certain risk; that is, if we accept a solution to a problem, and act as though the problem is solved in the pursuit of another end-in-view, then we run a risk of accomplishing this new goal less efficiently than we would have, had we accepted some other solution. So that, the pragmatist's criterion of resolution is not sufficient for evaluating our ways of deciding issues. One may *resolve* a problem in the sense that he no longer pursues the end that induced the inquiry, without *solving* the problem in the sense that the original end-in-view has been completely attained. Whether or not a goal or end has been attained must depend, on pragmatic grounds, on what instrument the goal will play with respect to more ultimate goals; and though we may adopt a certain response, and hence resolve a problem, the response adopted may not serve as the best instrument for our next goal.

Hence, to evaluate a given method of selection, according to pragmatic criteria, we must know for what purpose the selection is to be used. Hence, all "losses" involved in wrong risks are to be measured in terms of our purposes. See (2).

The examination of the relativist position with respect to inference leaves open a fundamental problem of his method. We must know, granting that the methodology of inference depends upon the means-end relationship, whether this relationship is a priori to all knowledge. Must we view the world through a particular form, a form determined by a priori teleological laws? How can we determine whether our purposes are actually being served by a given solution? Does the relativist agree with a critical philosophy and claim recognition of valid means-end relationships on intuitive grounds? If he does, he fares no better than his predecessors, and has failed to learn the lesson of history. No, the relativist must mean that we can set up experiments to test whether a goal is being effectively pursued. An experimental science of teleology is as necessary as a science of physics. But in the answering of questions

about goals, what are we to use as controls? Evidently, on pragmatic grounds, other ends-in-view. Are these arbitrarily selected, or are there criteria that enable us to decide whether our controls are adequate?

The *relativist* answer to this question is that the goals we use as a basis of investigation are always arbitrary; all inquiry is relative to such ends-in-view, and no absolutely verifiable answers exist. In effect, says the relativist, when we raise a question, we always raise it for an individual, or within some society or culture, and the answers we give are functions of the purposes of the individual or social groups. We may ask whether a decision serves such a purpose well or ill, but such a question can only be answered relative to other social purposes.

The reader will no doubt feel that the story of modern philosophy has taken us by a long and tortuous road back to an ancient philosophical position: scepticism. But the older scepticism and the newer are as different as night from day; the former was born of a dissatisfaction of *one* answer to the problem of truth (the Stoic empiricism), while the newer is the result of an examination of all historic answers. The older held no conviction even for its adherents, who refused to make *any* assertions about the problem of truth, but the newer relativism is certain on two points at least — namely, that questions of fact and questions of law are not absolutely answerable. Relativism can now say the only absolute certainty is that there is no certainty.

The role of scepticism is to show that there is no possible (consistent) theory of truth. This means that an exhaustive classification of schools has to be made, and a critical examination of each possibility effected. In such an examination we found it necessary to abandon rationalism for its failure to provide an adequate method for answering questions of law, and we found it necessary to abandon empiricism and criticism for their failure to provide an adequate method for answering questions of fact. Out of this examination, then, arose relativism, the position that factual and legal questions are unanswerable — a position we must apparently accept, since no other historic possibility remains.

However, we included in our list one other nonsceptical position,

a position that in its first reading seems impossible to maintain, for this position demanded of the scientist that in the answering of either questions of fact, or questions of law, he must always make presuppositions, and yet asserted that answers exist to at least some of these questions.

The experimentalist position seems to demand in effect that there be no *beginning* to the process of scientific method (i.e., that there be no questions answerable without presupposing other answers), yet that there should be an *end* to such a process (i.e., that certain questions be answerable in an absolute sense). Examples of processes or series that have no beginning but do have an end are not difficult to find in mathematics: e.g., the series of minus integers has an endpoint (-1), but no beginning, as does the series of real numbers greater than 0 but less than or equal to 1; hence, the experimentalist's postulates are not logically inconsistent since they have representations in the (supposedly consistent) field of arithmetic. But all such examples of a noninitiated, ending series are examples of series that contain an infinity of elements (in the sense that the number of elements exceeds any given amount). Hence, if we are to use such examples from mathematics in the proof of consistency, we must conclude that for the experimentalist those questions that are answerable should be answered after an infinity of steps in the process of learning. To make such a concept clear, we might say that the absolute answers are the limits of an infinite series of experiments or observations. By asserting that answers "exist" for some questions of fact and law, he means, using the mathematical analogy, that with an increase in the number of experiments it is possible to approach, in some sense, nearer than any given distance to the precise answer to a question asked.

If experimentalism is possible, then it must so formulate scientific method that it can give meaning to the term "precise answer," even though no precise answers are known, and give meaning to the operation of approaching this answer. In effect, experimentalism must define the *progress* of science in such a manner that (a) this progress has no starting point, (b) it has an end-point, (c) the end-point is "ideal" in the sense that it is the limit of an infinite series,

(d) the end-point is approachable in the sense that it can be determined whether science is nearer its ideal or not at two instants in its history. These are the difficulties and problems of experimentalism; the "value" of a research into these difficulties is that experimentalism remains the sole possibility of a nonrelativistic scientific method, i.e., of a scientific method that will provide a satisfactory answer to the problem, "What is truth?"

In order to examine the experimentalist viewpoint, we will first have to examine how science has gone about accomplishing its purposes, in order to determine what is nonrelativistic in this process. We must, in Dewey's terms, "turn to the historical development of science, in which is recorded what kind of operations have definitely been found to effect the transformation of the obscure and perplexing situations of experience into clear and resolved situations"[1] (6, p. 124).

We proceed in the next chapter to examine the manner in which science has attempted to accomplish one of the "simplest" of its problems, the answering of a single-valued question of fact. The sequel will show how this special aim can be generalized into a purpose for all science, a purpose that is nonrelativistic in character.

REFERENCES

1. Boring, E. G., *The Physical Dimensions of Consciousness*, Century (1933).
2. Churchman, C. W., and Ackoff, R. L., "Footnote to 'Logic of Statistical Tests,'" *Bulletin of the Institute of Experimental Method*, Vol. 1, No. 4 (1947).
3. Churchman, C. W., and Ackoff, R. L., *Psychologistics*, mimeographed, Philadelphia (1947).
4. Dewey, J., *Experience and Nature*, Open Court (1926).
5. Dewey, J., *Human Nature and Conduct*, Holt (1922).
6. Dewey, J., *The Quest for Certainty*, Minton (1929).
7. Freud, S., *The Basic Writings of Sigmund Freud*, translated and edited by A. A. Brill, Modern Library (1938).

[1] This, and many other passages in Dewey's writings, would seem to classify him as an experimentalist; but the fact that Dewey never explains in any explicit detail how these necessary tasks of a nonrelativistic pragmatism are to be carried out seems to leave him, if we judge pragmatically, as an *oral* experimentalist and an *actual* relativist.

8. Hegel, G. W. F., *The Phenomenology of Mind*, translated by J. B. Baillie, London, Swan Sonnenschein (1910).
9. Helmholtz, H., *Physiologiste Optik*, Leipzig, Voss (1896).
10. James, W., *Pragmatism*, Longmans, Green (1919).
11. James, W., *The Meaning of Truth*, Longmans, Green (1909).
12. Mill, J. S., *A System of Logic*, 8th ed., Longmans, Green (1872).
13. Peirce, C. S., *Collected Works*, edited by C. Hartshorne and P. Weiss, Harvard Univ. Press (1931).
14. Schiller, F. C. S., *Humanism*, Macmillan (1903).
15. Singer, E. A., Jr., *Mind as Behavior*, Adams (1924).

Chapter X Experimentalism I — The Answering of Questions

The dialectic of history which has been the subject of the preceding chapters has had as its purpose the answering of two fundamental problems that arose in connection with the methodology described in Chapter II: "Are presuppositions necessary for conducting an inquiry into any question science may raise?" and "Are there any answers to the questions of science?"

The tendency in modern science has been to give, as far as possible, an affirmative answer to both of these questions. Thus rationalism attempted to reduce the number of presuppositionless inquiries to those involving certain very general properties of the universe, while the empiricist attempted to reduce the number to certain very specialized aspects of an individual's mind (his sensations). Criticism and relativism, combined, succeeded in showing that the first question must receive an affirmative answer, if all aspects of inquiry are to be made intelligible, i.e., if uncontrolled intuition is to be removed from a perfect methodology of science. As long as intuition reigns we are not aware of all the presuppositions we have made, and the conclusions we reach are not verified in an absolute sense; they are only verified relative to the implicit presuppositions involved in our general or special intuitions. The remaining problem of methodology, the problem relativism alone raises, is whether or not answers to questions exist despite the necessity of presuppositions in inquiry. Our purpose now is to see in what manner we can define "verification" once we have accepted the necessity of presupposition.

In order adequately to refute the charge that relativism makes against any absolute answer to problems of science, it will be necessary for us to generalize upon the principal theme of the essay, and to attempt a characterization of experimental methodology. Only by thus determining what constitutes experimental inquiry in general can we hope to define that particular aspect of it which involves the concepts of probability and statistics.

We may begin our task with an intuitive or common sense understanding of what is the purpose of science. We feel that it requires no defense to say that one of the purposes of all scientific activity, *taken collectively*, is to reduce error to zero, i.e., to become absolutely precise. It is to be emphasized that the purpose is a collective one; distributively, the aim of this or that scientist may be something entirely different: to make a more efficient machine, to estimate the sensitivity of an organism, to predict the behavior of an individual, and so on. But a purpose that characterizes the sum total of all such activity, a purpose that makes the activity scientific at all, is the reduction of error. Whether there are other purposes, or whether this is the most general way to describe the purpose of science, are questions we shall postpone for later discussion.

We are thus in agreement with the relativist that meaning and truth in scientific inquiry are dependent upon the purpose of the inquirer, but we are going to deny the thesis that no absolute, conditioning purpose exists. Instead, we intend to see what happens to a relativistic purpose remodeled according to the absolute end of reducing error to zero.

In effect, we are postulating the relativist viewpoint out of existence for the moment; a thoroughgoing relativism could certainly raise the question whether the assignment of such a purpose to science is not an aspect of our present culture, and in general raise the question whether the meanings of the concepts involved in error-reduction are not relative to our own specialized ends. The only feasible way to escape such scepticism is to presuppose certain aspects of science, as a basis for constructing a nonrelativistic account of methodology. In the end, we shall have to return to the problem of our presuppositions, to determine whether they also can be considered in a nonrelativistic sense.

We begin the analysis by formalizing the manner in which science verifies (answers) its questions of fact, even though presuppositions are demanded at each stage of inquiry.

We take, as a typical example, a quantitative question, of the sort "What is the distance AB?" We say that this question will have been *answered* if there exists a uniquely defined real number k, such that the risk of accepting the hypothesis "AB is k units" is zero. Now to assert this much is already to have presupposed something: we have presupposed, in particular, a probability theory rich enough to define "hypothesis" and "risk." In effect, in order to begin to define the meaning of answer we have accepted the first step of the methodology of Chapter II: the formulation of a probability theory.

Our next demand is that the answering of the question "What is the distance AB?" presupposes a set of *pertinent* observations. We are now raising a problem that was left unanswered in Chapter II, but a problem that receives partial answering in the history we have been examining: by what criteria do we judge the pertinence of observations? The same problem can be formulated somewhat differently. We note that the question asked is ambiguous as it stands. What *is* AB? Is the distance to be taken at a specific moment of time? In other words, to judge the pertinence of a set of observations, we must specify the meaning of the question asked.

Now such specification can only take place by our constructing a formal pattern of nature of which we take the distance AB to be an event. An *example* of such a formalism would be the following:

1. A and B are points in a Euclidean three-dimensional closed space, S.

2. The *identification* of points in S is determined by principles of a Newtonian kinematics. (Hence, all points move on a continuous path relative to some fixed reference-system.)

3. The points A and B have a mass, besides a space-time individuation, and the laws governing their motion are the laws of a Newtonian mechanics (e.g., of Boscovitch).

By the use of such a model, we can re-phrase the question "How long is AB?" as follows: "What is the distance between the mass points A and B at time t_0 in the closed space S?"

Some such reformulation of the question is demanded, if the question asked is to have meaning in terms of observations, though evidently the particular formulation we have chosen is only one of many possibilities. Now just why an identification-principle, and the laws of motion, must be included will depend on a further analysis of the meaning of observation.

The history we have surveyed has shown that the meaning of pertinent observation must include the concept of sensation, so that the spatio-kinematical-mechanical image we have given is not sufficient to define pertinence. We must have a way of tying together the individual points of the formal construct with certain sensations of observers:

4. An operational correspondence exists between the entities in the formal construct and certain aspects of the sensations of an individual.

We have thus been led to include in the natural system observing minds; the formal manner in which we do this has been studied in some detail by the operationalists. The problem is to specify the operations a certain individual must perform, in order that a response on his part can be interpreted to mean "The distance AB at time t_i has been observed to be x_i." It will be noted that 4 actually includes concepts not definable within the limitations of a mechanics, e.g., "psychological responses." We will have to examine later just how such responses can be introduced without conflicting with the mechanical principles we must also presuppose.

Now the observations that are made by the specific operations given in 4 are relative to various times, t_i. If more than one observation is required, or if the observation cannot be made exactly at time t_0, we require a method of translating an observation made at one time into a pertinent datum for the answering of a question about another time. Such translation can be made by means of the kinematical and mechanical images given in 2 and 3, and the need for such translation is sufficient to establish the need for these images. As a consequence, we can now define the formal conditions under which an observation can be regarded as pertinent to the answering of the question "How long is AB at time t_0?"

It is to be understood that the formal conditions which now act

as the definition of the question do not have the "immutable" characteristic imposed upon them by rationalism and empiricism. It is certainly possible, and will eventually become necessary, to investigate their "validity" and to change their assertions.

Next, in order to "use" the pertinent observations in answering the question asked, we must also specify the statistical assumptions governing the selection of a sample, as the methodology of Chapter II indicates. These statistical assumptions might actually be deducible from the manner in which we describe the operations of making observations. That is, if the natural image is general enough, it will include sufficient information to characterize what kind of distribution function of the observations we can expect if the observations are made according to a certain set of operations.

An example of the application of statistical procedures to the present problem would be the following:

5. For any n, the set of the first n pertinent observations is a random sample from some normal universe, with mean μ and standard deviation σ.

In terms of the natural construct we have given, we can now make the following assertion:

6. The true distance AB at time t_0 is the mean μ of the normal universe from which the observations are drawn.

This assertion in general will be a consequence of the rest of the construct. Sometimes in practice, the true mean of a set of observations is not the true measure sought; e.g., when systematic errors occur in chemical analysis, the mean μ differs from the true percentage. However, a systematic correction of all observations can be made, such that the new mean and true percentage are identical. We suppose that such correction is involved in the complete description of the operations by which we make an observation. In other cases, the median, or a mode of the distribution, may be the true value sought.

Now the mean μ of a normal universe "exists" in the formal sense, since μ in this case is simply an integral of a single-valued function over the limits $-\infty$ to $+\infty$, and the integral can be shown to exist within the theory of limits. We might therefore be tempted to say that the answer to the question "How long is AB?" exists,

since μ and the "answer" are formally the same if nature is constructed as we have presupposed it to be. But such a definition is purely formal, and to regard such a definition of existence to be sufficient would be a characteristic of the school of rationalism. Since the answer exists only relative to a formal construct, we must inquire concerning the "validity" of this construct.

The criteria of the validity of the formal presuppositions, and of the "existence" of the answer to the factual question, depend upon the concept of a "stochastic limit." We say that a random sequence of numbers $x_1, x_2, \cdots, x_n, \cdots$, approaches a limit ξ *stochastically* if, given any ϵ and η, there exists an n, such that the probability P that x_n deviates from ξ by more than ϵ is $1 - \eta$. In less precise language, our assurance that the terms of the sequence become nearer and nearer their limit becomes greater and greater with increase in n. (See p. 106.) Such sequences, be it noted, differ in an important respect from the convergent sequences ordinarily studied in the theory of limits. For example, within formal theory, the sequence

$$s_1 + s_2 + \cdots + s_n + \cdots = \frac{1}{2} + \frac{3}{4} + \cdots + \frac{2^n - 1}{2^n} + \cdots$$

converges to 1 in the sense that for any ϵ, there exists an n such that

$$| 1 - s_i | < \epsilon, \text{ for all } i \geq n.$$

In general, stochastic sequences can oscillate considerably, and may have no well-defined characteristic. The nth term of a stochastic sequence may differ very widely from the limit, more widely, in fact, than any of the preceding terms; but the *probability* of divergences of a given magnitude decreases for an increase in the number of terms.

In terms of stochastic limits, and on the basis of the presuppositions made in 1 through 6, we can assert

7. The true answer to the question "How long is AB?" is the stochastic limit of the sequence $\bar{x}_1, \bar{x}_2, \bar{x}_3, \cdots, \bar{x}_n, \cdots$, where $\bar{x}_j$ is the mean of the first j pertinent observations.

The stochastic limit will exist, and will be exactly μ by 6, *if* we grant the formal presuppositions we have made. Hence, all the

formalism so far has only demonstrated the *conditional* existence of answers.

Now it is well known that any set of presuppositions such as we have given are subject to change within science; our examination of history has shown that Kant's project of finding invariant presuppositions that will be satisfactory for all observations must be abandoned. Hence, the conditions under which we have demonstrated the existence of answers to questions do not hold in any special case within the history of science.

There still remains the question whether or not at any stage of science it is possible to find *some* natural construct which enables us to accept the hypothesis that the sequence of means is approaching a limit stochastically; and in general, there is the question whether or not *some* ideal natural construct exists, though as yet unknown, which will satisfy for all possible stages of science the conditions for the existence of a stochastic limit. If such a construct does exist, then we can say that answers to both factual and legal questions exist within science: the answers to factual questions exist in the sense that they can be approached stochastically, and the stochastic limit exists; the answers to legal questions exist, in the sense that the formal construct within which the limits exist will be the "true" construct of the natural world.

The fundamental postulate of experimentalism, therefore, is the following:

There exists a formalization of nature, such that stochastic limits exist for certain sequences of mathematical functions of the observations which are pertinent to a given question of fact.

Both the "true" laws of nature and the "true" facts are "ideals" in the sense that for any finite *n* they are only approachable, never attainable. They are also "real" in the sense that they completely explain the real world. Idealism and realism are thus resolved within a complete scientific methodology.

We have said that the principle is a "postulate" of experimentalism, but we do not mean to imply thereby that this postulate is an arbitrary formal assumption. In fact, we shall want to show why its various formalizations are well motivated within our culture.

For the present our task is a more specialized one: to determine

exactly how we are led to abandon one type of presupposition in favor of another in the process of answering a question. For this purpose, the techniques of statistics are applicable, provided we specify what alternative presuppositions there are, and provided we specify the method of abandoning one set of presuppositions in favor of another. In effect, this aspect of scientific method is a check on the correctness of supposing that the pertinent observations are approaching a limit stochastically, and this is a check on the validity of one or more aspects of the formal construct.

A typical example of such a procedure would be the following: we decide to test the validity of presupposition 5, that the observations are for any n a random sample from some normal universe. The alternative hypothesis to 5 we take to be the following:

5*. The set of pertinent observations consists of subsets of equal size, N; each subset is a random sample from a normal universe with mean μ_i and standard deviation σ, where the μ_i are not all equal.

This kind of test might arise where the observations are made by different observers, each observer making the same number of observations; or it might arise in the case where a certain number of observations are made on a series of days. The experimenter then wishes to determine whether all the observers are equally reliable, or whether the same conditions applied over all the days. One formal construct (5) would assert that there was no significant effect in the results due to observers (or due to days), while the other, 5*, would assert that some account must be taken of the difference in observers (or days).

Various tests are available in the statistical literature for deciding which of the hypotheses should be selected on the basis of the data.[1] It is to be noted that the above is by no means the only test of control of observation that could be run. For example, another alternative to 5 might be that there is a gradual downward trend in the magnitude of the observations over a specified set of them. This hypothesis would be tested by a serial correlation coefficient of some sort. Or we might wish to test whether the closed space was Euclidean, and propose instead a non-Euclidean construct

[1] See references at the end of Chapter II.

which would modify in some way the pertinent observations. Such modification, for example, might result in a different hypothetical mean of the universe of the observations. The revised construct would then be tested by selecting one of the following two hypotheses: H_1, the observations up to any n are random samples from some normal universe with mean μ_1 and standard deviation σ, and H_2, the observations up to any n are random samples from some normal universe with mean μ_2 ($\mu_1 \neq \mu_2$) and standard deviation σ.

Now suppose that 5 is abandoned at some stage in the sequence of observations. Then what procedure is the experimenter to follow? There are actually a number of reasons why 5 should fail: The true natural image may not be Euclidean in geometry, or Newtonian in kinematics or mechanics, or the operations for collecting data may not be pertinent, or the assumption of randomness or normality may not be correct.

At this stage, the experimenter is faced with a choice. He must decide in what manner the natural image is to be reconstructed so that a reformulation of 5 is not discarded by the statistical test. Such moments in the history of science are well known: the Michelson-Morley experiments on the velocity of light, irregularities in the orbit of the planet Mercury, Bessel's discovery of the personal equation in astronomical observations, are but a few. In fact, in ordinary laboratory research this occurrence is a common item in the day's work: an experiment is designed and attempts are made to control all possible factors. The resulting data, however, are "erratic," as judged either by well-formulated criteria or by intuition based on experience. The experimenter has then to decide which of his presuppositions he must modify in order to overcome the difficulty: he checks his balance, calibrates his thermometers, and so on.

In effect, when we have been led to abandon hypotheses similar to 5, we are led to the conclusion that the observations are not all measures of the same quantity, and hence that the true answer is not being approximated according to any known criterion. The experimental task now becomes one of investigating the presuppositions we have previously made. At the present stage of science, there is no well-formulated technique for examining the

"suspects" when experimental control breaks down, though there seems to be no reason why such techniques could not be developed. Usually, the hint as to the origin of the trouble will be found in the manner in which the data are erratic, and the experimenter can often predict where his presuppositions have been faulty by examining the way in which the observations deviate from expected values.

If a given source of trouble is suspected, then one of two courses are open. The experimenter may simply assume an alternative presupposition, and proceed to make observations on the same quantity as before, or he may actually set up a new experiment to determine which presupposition is best confirmed by observation. In either case, the validity of a presupposition is the object of research, and if a new experiment is devised, then a new procedure similar to that discussed above must be used.

An example will assist in showing how the experimental situation may turn back in its progress and examine the validity of the presuppositions it has been making. A metallurgist is investigating the hardness of a metal with respect to its indentation on impact. Now the amount of indentation depends (at least) upon the energy of impact and the radius of the indenting pin, so that the question of "hardness" is ambiguous without further specification. Again, in order to specify the meaning of the question asked, the experimenter must identify the type of material under investigation, i.e., he must state the conditions under which observations of indents are pertinent. For example, suppose he makes the following presuppositions:

1. The identification of the alloy under investigation depends upon certain chemical analyses and grain-size readings (as specified by a range of values).

2. Pertinent observations are to be made by using a pin of a certain type of steel and certain radius of point, by holding the pin in a certain manner over the alloy, when the latter is placed on a flat steel anvil, by striking the pin with free-falling weights at various speeds, by measuring the depth of the resulting indent below the highest point on the indented surface. (All of these specifications, be it noted, are in terms of ranges of values; for example, the radius of the pin point may be specified as 0.0370″ ± .0005″.)

3. When the data on indent depth collected in 2 are plotted against the energy of impact of blow, for various known energies, they deviate in a random fashion, according to a normal law of error, from some straight line, provided the velocity of impact exceeds a certain v_0, and is less than a certain v_{max}.

4. With an increase in the number of observations, the least-squares estimates of the slope and intercept of the straight line in 3 approach stochastically the true slope and intercept, and the stochastic limits are defined as the "real" measures of hardness of the metal.

Now an experimenter who follows out the above procedures may begin to make pertinent observations, but he will soon find that the slopes estimated over one range of energies differ significantly from the slopes estimated over another range. Suppose he sets up the following problem of experimental control:

H_1: accepts 1, 2, 3, and 4.

H_2: accepts 1 and 2, and asserts 3*: the data in 2 deviate in a random fashion, according to a normal law of error, from some ith degree line ($i > 1$).

Then in general he will be led to abandon H_1, and hence to abandon one of his presuppositions. Now he need not abandon 3 as it stands. He may argue that the nonvalidity is due to factors he has failed to include in 2 (e.g., to failure to control temperature), or to inadequate identification-specifications in 1. Actually, it has been found that 3 is incorrect as it stands; instead of the *depth* of indent, the *volume* of indent must be calculated in 2, and the resulting data will apparently (at our present stage of research) deviate normally around a straight line; the slope and intercept of this line may then be defined as the true measures of hardness.

The experimental process we have been describing is usually called the process of experimental control, and our task has been to formalize the nature of such control in order to describe science's way of approaching its ideal of absolute precision. To summarize, *an experiment is said to be controlled if we state all the formal conditions under which a mathematical function of a series of observations approaches a limit stochastically.*

The definition of experimental control we now make the criterion

of meaning: *No question of fact can be said to have meaning unless there exists a controlled experiment for its answering.*

We shall want to expand upon the meaning of control and the meaning of meaning in the subsequent chapters. We shall also want to motivate the choice of the experimentalist postulate instead of a relativist analysis of methodology. For the moment, we present a summary of the analysis of experimental method so far made.

The answering of a single-valued question, we have said, demands, in addition to a series of observations, a formal image of nature that enables one to make adjustments of observations. Neither observation nor theory can be considered as enjoying "certainty" at any finite state in the series; but methods exist for determining whether the observations are in control, and whether the formal theory is adequate. Further, if the postulate of experimentalism is granted, it is always possible to find an image that will enable us to reduce the "error," with an increase in the number of observations, to a quantity less than any given amount. The degree of precision of a set of observations is also measurable; one way of viewing this is to say that the question "What is the (true) magnitude of the error for any given set of observations" is itself a single-valued question, susceptible to the same methodology as that already described (an "error of the error" can be measured, and reduced indefinitely). In terms of the basic methodology of experimental science we can then define the concepts that are fundamental to any theory of knowledge, meaning, truth, and reality:

1. No question of fact can be said to have meaning unless its answer is a stochastic limit of an infinite series of approximations.

2. No question of law can be said to have meaning unless it forms a part of an image of nature which can be used as a criterion for the adequacy of a set of observations.

3. The true answer to a question of fact is that single value for which the error of observation is zero.

4. The true image of nature is that image which will produce experimental control for all series of observations, finite or infinite.

From the Postulates it follows that

Theorem 1. The method of answering any question of fact presupposes the answering of questions of law (since the true answer to questions of fact requires the true image of nature).

Theorem 2. Answers to questions of fact *exist,* in the sense that they are stochastic limits of an infinite series of approximations.

Theorem 3. The method of answering any question of law presupposes the answering of questions of fact (since the definition of a true image of nature depends upon a true answer to factual questions).

Theorem 4. Answers to questions of law *exist,* in the sense that they are limits of a series of approximations designed to give better and better control.

These theorems are taken to be those characterizing the philosophical position of experimentalism in Chapter IV.

REFERENCES

1. Churchman, C. W., "Probability Theory, I, II, III," *Philosophy of Science* (1945).
2. Smith, H. B., "Postulates of Empirical Thought," *Essays in Honor of E. A. Singer, Jr.*, Univ. of Pennsylvania Press (1942).
3. Singer, E. A., Jr., *Experience and Reflection* (to appear).
4. Singer, E. A., Jr., *Mind as Behavior,* Adams (1924).
5. Singer, E. A., Jr., "Philosophy of Experiment," *Symposium* (1930).

Chapter XI Experimentalism II — On Meaning and Method

In the last chapter we have outlined an example of the manner in which the scientist answers questions, according to the most general criteria of answering. The purpose here is to develop the basic methodology exemplified in the last chapter; this methodology in effect is a generalization of the methods of Chapter II.

We shall first state three principles which we take to be at least the necessary conditions for the meaningfulness of any question. The subsequent discussion will be devoted to the general aspects of these principles:

A. Formal conditions must be given which specify the exact manner in which pertinent observations are to be made.

B. Formal conditions must be given which enable one to make a *response* to the question asked on the basis of a finite set of observations.

C. Formal conditions must be given which guarantee an indefinite approximation to the *answer* to the question asked (in the case of factual questions, in the sense that the true answer is a stochastic limit).

These three conditions actually specify the manner in which questions are to be framed within an adequate methodology of science. The pattern outlined in Chapter II is correct, except that the general problem of question-answering demands far more than is supplied by statistical presuppositions alone. If a question is to be phrased in a precise manner, we will require a set of alternative hypotheses, as in Chapter II. But now the alternatives must include

presuppositions which state the conditions under which an observation is to be regarded as "pertinent" (principle A), and the conditions under which an indefinite approximation can be made to the true answer, i.e., to an answer given without risk (principle C).

From the viewpoint of this trichotomy, we can think of the precisely phrased question as made up of three parts. Each alternative answer will include a set of assertions which enable us to distinguish between pertinent and nonpertinent observations, a set of assertions which enable us to make inferences (subject to risks) on the basis of a finite set of observations, and a set of assertions showing how an indefinite approximation to a risk-less answer can be given. As in Chapter II, those assertions which are common to all the alternative hypotheses we call the "presuppositions of the inquiry," while those that differ we call the "basis of inquiry." Of course, in actual practice, this trichotomy need not be followed strictly, since it may be simpler to state certain principles under one heading which will imply all the principles needed under another heading (for example, sufficiently inclusive statements about the manner in which the true answer can be approximated might supply all the necessary statistical assertions for the satisfaction of principle B).

It is to be noted that principle B, while its meaning includes the methods given in Chapter II, is to be viewed in a more general manner. The principle as we have stated it includes the term *response*, as distinct from *answer*. We say that an individual, or group of individuals, gives a *response* to a question, when he acts as though a certain answer were true. This means that a response is determinable by the behavior pattern of an individual. But not every behavior pattern is a response; we must add the further condition that the behavior pattern is a *means* in the accomplishment of an end, the meaning of these teleological concepts to be explained in the next chapter. As we shall later argue, not only can we determine by experimental methods the response an individual gives to a question, but we can also measure the *efficiency* of this response. Thus, to take a trivial example, an individual who responds as though there were furniture located at various spots in a room will exhibit a certain type of behavior, and this behavior will be efficient

if it does not defeat the goal, say, of passing through the room in a certain time interval. The responses we give to questions, therefore, are not merely the basis for action; they *are* the action itself. The inspector who passes a lot of material is acting as though a certain answer to a question were true; the efficiency of this response will be measurable in terms of the efficiency of his behavior in this instance relative to the uses to which the manufactured article is put.

Hence, principle *B* may be taken to have the following meaning: Conditions must be given which enable us, at the end of any number of observations, to give a *response*, i.e., to decide which course of action will be most efficient in accomplishing a specific purpose. The defining purpose of a question must be made explicit enough so that at any stage we can decide the best action which can be undertaken towards accomplishing the purpose. Suppose I wish to know the way in which certain chemicals should be combined to make a tooth powder. According to this principle, I must so specify my purpose in raising this question that I can decide, on the basis of a finite set of pertinent observations, what the best course of action will be, i.e., which combination will be most efficient in cleansing teeth to a specified degree of purity.

What we have called in Chapter II a "method of selection" is therefore embodied in the assertions demanded by principle *B*. The principle demands that our assertions be rich enough to enable us to decide which course of action is the most efficient on the basis of the observations.

It is the motivation for principle *C*, the demand for an indefinite progress towards the "true" answer, that needs most defense here. The relativist might be characterized as arguing that we can only give responses to questions, not answers. He can see no need of adding an idealist principle to *A* and *B*, which together supply all the demands for methodology of experiment. The most general reply to such a relativist position we shall reserve for the later discussion of the general meaning of "efficiency." For the present, we may point out the reasons that exist within science itself for insisting upon principle *C* as a necessary condition for meaning.

These reasons lie in the possibility of investigating the presuppositions of inquiry. Our historical survey has shown that we can

find no set of assertions which will guarantee (without risk) the pertinence of an observation; we can find no set of general principles which will guarantee general aspects of the natural world. If this is so, then the manner in which principle B is satisfied, i.e., the techniques by which we decide the most efficient course of action, are relative to the accuracy of our presuppositions. In general, we must be able to raise, and to investigate experimentally, such questions as whether the conditions specifying pertinence really do so, whether the conditions really are stated precisely, whether the method of selecting one of the alternative hypotheses does actually present the most efficient technique. These investigations of the presuppositions of inquiry can only have meaning if we assume that there is some criterion of efficiency, and in general, if we presuppose that a most efficient method exists. But the most efficient response to a question is exactly what we mean by an "answer," for the most efficient possible response will be a behavior pattern which is sure to attain its goal, if the goal is attainable. That is, the risk entailed in a perfect response is zero. It is only in terms of such a standard of absolute efficiency that we can investigate the adequacy or inadequacy of our experimental techniques, i.e., of our presuppositions. Hence, *in order for a response to have meaning, we must presuppose the existence of an answer.*

Thus science can study the typical responses individuals and social groups make to questions; it can also study the efficiency of these responses relative to specific goals. And if the responses have been made consciously, so that the responder has made explicit the presuppositions of his inquiry, science can investigate the validity of the presuppositions. The operational viewpoint we have already discussed is correct in its demand for presuppositions which describe the manner in which pertinent observations are to be made; its viewpoint is *empirical*, in its neglect to specify the manner in which these "controls" are themselves to be investigated. We next say that the validity of the operations are dependent upon the purposes of the inquiry. This much alone is *relativistic*. We say, further, that the measure of efficiency of a response (and hence its presuppositions) depends upon a standard of absolute efficiency, and this much is *experimental*.

Now if we define the *purpose* of science to be the attainment of answers to questions, in the sense described in the last chapter, then there arises a peculiar feature of the response-problem which certainly demands our attention here.

Science herself is defined by a purpose, and she exhibits in the behavior patterns of the experimenter various "responses" to her own typical questions. Science's characteristic ends are *ideal* in the sense that they are unattainable but presumably indefinitely approachable. These are the *characteristic* ends of science, be it noted; they are characteristic in the sense that he who pursues them is thus far scientific. He may have other goals or ideals as well. The medical research worker who discovers a cure for a disease is scientific in so far as his discovery advances science one step further on the road to an absolutely precise answer to a question; he is also, of course, pursuing another end (say, longevity). The "advantage" of pursuing the ideal of science is that as a consequence we at the same time pursue all our other ends more efficiently.

Now the peculiar feature of the scientific situation is this: Can science study its own responses? Can it evaluate these responses? This feature of science has seemed so peculiar to some that they have used it as a basis for segregating philosophy and science. For philosophy, having divorced itself throughout the ages from one after another of the special disciplines, has finally assumed the role, in modern times, of the grand evaluator of human activity. The philosophy of science is thus supposed to play the role of the evaluator of the responses scientists give to questions. It plays this role by deciding what are the ideals (ends) of science, and by deciding whether or not a certain method is adequate to this ideal.

The dangers of such a segregation of philosophy and science are clear enough. We then have a situation in which one mode of inquiry (science) depends upon another (philosophy) for the criterion of evaluation, but not conversely. The result is that philosophy is able to slip away from any controls into assertions unchecked by outside criticism. The philosophy of science therefore becomes rationalistic in its methodology, even though it abandons rationalism as an adequate method of inquiry for science.

The alternative is that science can evaluate its own responses;

that the philosophy of science *can* be made a science of philosophy. The task of such a science is twofold: one, to determine the ideal of science by scientific methods; the other, to describe the manner in which science can most efficiently approach its ideals.

In order to see in what way a science of efficiency can be developed, we must note a certain differentiation between purposes: that some purposes are relative to individuals, some to social groups, some to a culture, and some to a history of cultures. Hence, when we attempt to determine what a certain question "means," our inquiry is apt to be ambiguous until we specify the origin of the purpose that lies in back of the question-asking. The subsequent chapter on personality and social groups will try to show the different characters associated with problem-solving on the individual and social levels. The argument contained therein will attempt to show that (a) the determination of purpose, whether of individual or social group, implies a criterion of efficiency, and (b) that the meaning of this criterion cannot be found within either a study of the individual or the society. This implies that the meaning of purposes, and hence the meaning of questions in general, presuppose principles of a science that lie "beyond" the science of social groups, a science, as we shall argue, whose material is the history of societies.

The concept of efficiency will not be the only one in science whose meaning will depend upon its history within many cultures; the concepts of "mass," "time," "life," "mind," "personality," "intelligence," "social group," would be but a few that share this characteristic.

This brief discussion can now be made the basis for the following general account of the manner in which the analyst should bring history to bear on the investigation of the presuppositions of an inquiry, or, what is the same, the manner in which the analyst should use history to investigate the meaning of a question. The term "history" is here taken to be synonymous with organized experience.

1. He must formulate alternative hypotheses concerning the meaning of the question asked, i.e., concerning the historical purpose in raising the issue.

2. Each alternative must be complete enough so that the information of the historian can be applied in the selection of one of the

alternatives; specifically, each alternative definition should comply with the three principles A, B, and C, given above.

3. His study should be relative to the use of certain language symbols, though the historical study of language will be by no means *sufficient* for the selection of one of the alternatives. If he feels that the same word has two or more meanings, i.e., if he feels that questions raised about the same word have quite distinct purposes, then the historical information should be employed to show that such an hypothesis is justified. For example, present-day disputes as to the meaning of "probability" might be considerably clarified if a well-designed historical "experiment" on the meaning of this concept could be made.

In this sense, the history of science, or past experience, becomes a fundamental part of all methodology. It is the basis of inquiry into the legitimacy of a certain method of defining. Such a history, if properly designed, should enable us to decide the fundamental purposes man has had in raising certain issues; it should enable us to determine whether certain techniques for answering a question are "adequate," i.e., whether they constitute the best means for accomplishing the historical purpose.

An example of the manner in which such a history would be of assistance within methodology will be helpful. We make reference to the example of the previous chapter. Let us suppose that the experimenter investigates whether or not his observations on the length AB are in experimental control. His investigation, let us say, is based on the hypothesis that there is a "trend" in the data he is collecting; he tests this hypothesis by the use of a serial correlation coefficient, and selects the hypothesis that trend exists, and hence abandons the hypothesis that experimental control has been maintained. We have said that this amounts to abandoning one or more of the presuppositions of inquiry. As Singer has put it, a set of presuppositions are said to be "permissible" relative to a set of observations if the hypothesis of experimental control is not abandoned under one of its many forms. This procedure raises the following methodological issues (among others):

1. What hypotheses of lack of control should the experimenter raise? (E.g., should he raise questions about randomness, or the

characteristics of the universe, or the legitimacy of procedures in collecting certain of the observations?)

2. If the hypothesis of lack of control has been accepted, then what modification in the presuppositions should be made?

In the case of the second question, the choice of presupposition regarding the natural world, Poincaré has suggested the criterion of "simplicity and convenience." His suggestion is illuminating from the point of view of the present discussion, for simplicity and convenience are terms we normally ascribe to means for the purpose of accomplishing certain ends. If the proposed science of history could be carried far enough, we should be able to evaluate from the point of view of efficiency the responses an experimenter gives to either of the above questions. In some cases, it will be more efficient *not* to raise the question of control. This seems to be especially true in the earlier phases of research, where it is important to collect a lot of information, even if a good deal of it is "unreliable," in order to formulate more exactly the procedures for later experimentation. On the other hand, there will be cases where a great deal of time and effort is lost in collecting data which are useless, i.e., useless from the point of view of the fundamental purpose of the inquiry. Again, it may be preferable to use a very simple natural construct, even though we are prevented thereby from carrying our investigation into certain fields, simply because such a construct will afford the most efficient way of proceeding with the inquiry. In some cases, perhaps most, we may find it methodologically profitable to keep contradictory tenets within science. Logical consistency has no necessary priority.

Although we have only sketched in very broad outline the kind of history of science demanded by methodology, and the general manner in which a study of meaning is effectuated, nevertheless we can characterize the differences between this viewpoint and certain other attitudes towards defining. For the rationalist temperament, there are certain invariant aspects of defining: these are concerned with the immutable "ideas," usually taken to be innate. An adequate definition is therefore judged in accordance with the fundamental concepts of innate reason. For the empiricist, on the other hand, the meaning of a concept depends upon its reduction to the

fundamental direct observations, so that adequate defining depends upon stating a specific set of operations which will supply the necessary data to answer a given question about the concept. The hardness of a metal for the empirical operationalist is therefore defined by a set of observational operations, stating exactly how indent measurements are to be made and the manner in which the indents are to be measured. Such a set of empirical operations completely specify the meaning of the term, according to empirical methodology.[1] To the relativist, the fault with either an empirical or rational analysis of meaning lies in its attempt to fix *one* means for an end. The means-end relationship (as later defining will try to show) always implies that no means is a *necessary* condition for the attainment of the end. Consequently, to ascribe one set of principles, or one set of empirical operations, as the basis of definition is to ignore the purposefulness involved in question-asking. It is a position comparable to the intuitionists in ethics, who attempt to ascribe one set of fixed principles defining the moral act; indeed, the history of ethics, like the history of methodology, may be viewed as an attempt to define the concept of "value," and with the development of relativistic and experimentalist philosophies, we are able to approach the formulation of an experimental science of ethics. The experimentalist is in agreement with the relativist that there is no set of invariant principles or operations which specify the meaning of a concept. He differs, however, in that the purpose fundamental to the definition must be sought within a broader scope than the individual or his society. See (1).

All these remarks on methodology and meaning have shown that the concept of purpose is basic to the experimentalist analysis of science. To make this analysis clear, the next task is to determine as specifically as possible the meaning of "purpose." We take this analysis to be based on the assumption that whatever may be the meaning of purpose, this meaning must be consistent with the physicist's aims, i.e., it must not conflict with a physical interpretation of nature in accordance with deterministic laws. It is to such

[1] The use of "indicators" (2) is merely a refinement of the empirical demand for reduction-sentences; its use implies the viewpoint that the meaning of a concept depends basically upon empirical evidence.

an experimental definition of purpose that we turn in the next chapter.

REFERENCES

1. Churchman, C. W., and Ackoff, R. L., "Varieties of Unification," *Philosophy of Science* (1946).
2. Kaplan, A., "Definition and Specification of Meaning," *Journal of Philosophy* (1946).

Chapter XII Experimentalism III — Non-mechanical Concepts

The reader will be a patient one indeed who has not long since raised in his mind what appear to be pertinent and pressing problems concerning the description of experimental method that has so far been given. These questions are mainly concerned with matters of omission: for example, we have said that if a lack of control is shown, then the presuppositions should be changed. But how changed? What principle or method guides the selection of a new image of nature when the old is no longer permissible? Or again, we have said that "observations should be made." But what is an observation? What is the experimental definition of an observer, and how do we study his activities?

We have on the one hand made a demand for a determinism in nature, in order that the true answer may be approached indefinitely, and on the other hand we have described the process of question-asking to be purposive. How are purpose and determinism reconciled?

All these questions we assume to rest upon the same kind of omission in the description we have given. The direction to be taken in making the postulates of experimental method *sufficient* can best be seen by another example. It is all too frequently emphasized that the exactness and precision of formal theory can never represent the data of our senses; in the sense world, for example, no "line" is ever straight, no physical circle ever satisfies the criteria of a formal geometrical circle. It is sometimes inferred that the world of theory and world of sense cannot be made to meet on a common

ground; the problem of "application" is therefore supposed to present insurmountable difficulties to the complete methodologist, and we are purported always to require a certain amount of intuition in the application of the theory to sense objects. If this problem of application can be translated at all into the terminology we have introduced up to this point, it must mean that the uniqueness demanded of elements in our formal image of nature can never be enjoyed by the objects of observation. The problem of correspondence of these formal elements and the objects of observation can therefore never be made exact at any finite stage: the observed may correspond to any one or any set of elements in the image.

But this does not force us to the undesirable position of asserting that we cannot become more methodical, more precise, in the required correspondence; the problem of application is only an insurmountable problem provided we regard observation (sense data) as completely independent of theory. If we can enlarge our image of nature in some sense, so that it includes the lack of precision to be found in the nature of observation, so that, in fact, it includes a definition and theory of observation, then there is no reason why we should not become more and more precise in the problem of application. In effect, we regard questions concerning the properties of an observation to be questions as meaningful as questions concerning the properties of a certain body, and to be subject to much the same methodology in their answering. In the latter case, we have already given some account of the formal image that will be required to approach an answer; we have hitherto failed to do so for the case of questions about sensations or observations. Similar remarks can be made about the other problems mentioned above; we must so enlarge our image of nature, and so extend the methodology of experimental science, that we can state problems concerning the manner in which individuals formalize or theorize in a meaningful manner.

There appears to be a difficulty, it is true, in the proposed extension of the image of nature for the purpose of answering questions of this type; the sort of determinism we have demanded of our image seems to preclude any additions, other than additions to the number of elements or their dimensions. The problem of the extension of

such images is the basic problem of the mechanist-vitalist controversy; the mechanist asserts that the elements and properties of mechanics provide a sufficiently extensive image of all nature; the vitalist asserts that nonmechanical "forces" or individuals must be added to the mechanical image. The *logical* solution of this conflict may be found in a complete form in a series of papers by E. A. Singer (4), (5), and (6).

In brief, the extension of the image is accomplished by the introduction of a classification of classes of elements in the mechanical image. We say that two manifolds of mechanical elements belong (a) to the same *mechanical* class if a one-one configurational equivalence can be found such that the corresponding elements have identical structural properties, (b) to the same *physical* class if there is a function of the structural properties such that the two manifolds have the same value of this function, and (c) to the same *morphological* class if there exists a mathematical function of the properties which does not differ by more than k between the two manifolds. Morphological classes therefore represent ranges of values in the formal image. With respect to these classes, one may have two interests: one may wish to determine relationships between various distinct (exclusive) classes, or one may be interested in the dispersion of a given property within any class. This concept corresponds closely to the statistician's within-group and between-group variance, and is in a sense a generalization of this concept.

On the basis of the concept of morphological classes, one can make a formal distinction that is essential to an adequate characterization of probability. This is the distinction between the relation of "cause and effect," and the relation of "producer and product." To define the relationship of "cause" within the formal image of nature, we suppose that there is some dimension of this image for any value of which we can describe the "state" of the elements of the image; i.e., for a fixed value of this dimension, we can specify the dimensional values of all the elements. Under the classical mechanics, the dimension is "time," and any moment of time gives us a "cross-section" of the state of nature; under the special or general theory of relativity, the time-dimension would have to be given a point of reference in order to obtain unique states. In any event, because

of the determinacy necessarily operating in the natural image, one can say that each "cross-section" of the elements as defined by the reference dimension ("time") is both necessary and sufficient for any other cross-section, in the sense that the formal laws of the system, plus the elemental properties of one cross-section, are both necessary and sufficient to give the elemental properties of any other cross-section. We might phrase this by saying that chance fluctuations, or probabilities, do not operate in the cause-effect relationship, but we shall rather want to use this phrasing as a basis of defining "chance" and "probability" in formal terms.

The relationship of producer to product is defined in terms of morphological classes; we say that one element x (e.g., an acorn) of a morphological class X is the producer of an element y (an oak) of a morphological class Y if the following conditions hold:

1) x and y belong to different cross-sections of the same natural image.
2) x is a necessary condition for y (had there not existed this acorn in a specified region, then no oak would have existed in this region at a later time).
3) x is not a sufficient condition for y (it does not follow from the laws of the natural system that if the oak had failed to occur in its region, the acorn would not have existed).
4) there exist other members of the x- and y-morphologies which belong to different cross-sections of the same natural image but fail to satisfy 2) or else fail to satisfy 3).

In less formal terms, x produces y if it is essential to y's existence, but not the only essential thing. Thus the father produces the son, and is essential in this production, but evidently not sufficient; further, cases of nonproduction of the same morphology (human) are common enough.

Certain important differences should be noted between the cause-effect relationship, and the producer-product. In effect, common usage is ambiguous with respect to the causal relationship; on the one hand, "A causes B" is often meant to imply that if the condition A is given, B necessarily follows, and on the other hand "A causes B" is also meant to imply that had not A occurred, then B

would not have occurred either. In terms of *occurrence at a specific time*, there are four possibilities (granting that A precedes B in time):

1. The occurrence of A is *necessary* but not *sufficient* for the occurrence of B: If not-A, then not-B, but if A, then not necessarily B.
2. The occurrence of A is *sufficient* but not *necessary* for B: If A, then B, but if not-A, then not necessarily not-B.
3. The occurrence of A is both necessary and sufficient for B.
4. The occurrence of A is neither necessary nor sufficient for B.

In common usage, any one of the first three possibilities is designated by "A causes B." To remove ambiguities, we have restricted the causal relationship to 3. This means that the *domain* of the relationship "is the cause of" is restricted to "cross-sections" or "time-slices" of nature, at least within a natural system of rigid connections, or its analogue. For the only aspects of such systems that can be said to be both necessary and sufficient for their respective occurrences, are the complete time-slices. If only a partial aspect of two time-slices is considered, then changes in the remainder of the slices can be made so that the occurrences will not take place, i.e., A will not be sufficient for B.

Since the domain of causality consists of time-slices, then we cannot talk of one individual *causing* another to exist. This aspect of common usage is represented by the first possibility, where A is a *necessary* but not sufficient condition for B. When this aspect has been exactly formalized, we say that A "produces" B. The exact formalism entails (1) a description of A and B in terms of morphology; (2) an individuation of A and B in regions in their respective time-slices; (3) the establishment of A as necessary for B, in the sense that failure of something of the A-morphology to occur in the A-region implies by the laws of the mechanical image the failure of something of the B-morphology to occur later in the B-region; (4) the establishment of A as nonsufficient, in the sense that the absence of something of the B-morphology in the B-region does not imply that earlier there was no A in the A-region. It might be noted that condition (2) makes B a "producer" of A; i.e., a future event that determines a past.

The domain of the relation "producer" is, therefore, individual sets of mechanical points classified morphologically. Thus, while we speak of entire time-slices of a closed system *causing* other complete time-slices, we talk of individuals, i.e., partial aspects of time-slices, *producing* other partial aspects. Under this terminology, it is improper to say that the hammer *caused* the nail to move forward, for there were other aspects of the time-slice in which the hammer occurred which were also essential to the final result; but we can say that the hammer (along with many other things) was the *producer* of the given result. We say that the remainder of the time-slice of a closed system, other than the individual under consideration, is the *environment*. A re-examination of the conditions of the producer-product relation will show that the environment is also a producer, if the individual set of points is. That is, the individual and its environment are "coproducers" of the object.

By certain generalizations, the formalism of which we shall not consider here, it is possible to regard the *behavior pattern* of an individual over a segment of the time-continuum as a producer of either an object or another behavior pattern. This generalization requires not only a morphological classification of individuals, but also of behavior patterns of the mechanical points that compose the individual.

The next step in the analysis consists in introducing the concept of "function." Let us first consider a simple example. We say that an individual wrist-watch produces a certain response in certain people, a response that we can designate as "telling-time." This response can be described by a morphological definition of behavior patterns among a class of individuals. Now we also note that the sun-dial, the clock on the mantle, the chronometer, also exhibit the same producer relation. Suppose, now, we form a class of all such instances of production. It will be clear that the assertion that an object belongs to the class of such producers does not depend for its validity upon the structure of the object itself, for there is no common morphology to be found among timepieces. We say that the class of all such producers is a "functional" class, and that the property which defines the class is the "function" of each of the members. In a sense, then, we have introduced three separate ways of "explaining"

an individual's behavior. We may regard the behavior by itself, as obeying certain mechanical laws; each mass-point of the individual will follow a path prescribed by the differential equation of the natural image and the boundary conditions. Second, we may consider this behavior pattern as an instance of a certain producer-relation, by classifying the behavior morphologically, and determining whether the producer-conditions are satisfied. Third, we may determine whether the present instance of production may be classified within a similar set of instances, where the defining property of the set is independent of the morphologies of the elements entering into the production-relation. The falling of an axe may be described with respect to its motion by the mechanical laws of falling bodies; it may also be described as producing a split in the log. It may finally be described as having the function of log-splitting, since many other morphologically dissimilar behaviors may produce the same result.

Finally, if a set of behavior patterns of the *same* individual can be classified by a criterion that is independent of morphological properties, then the individual is said to exhibit a "purpose." For example, the complicated behavior patterns of a man during a day may be classified under a common property, i.e., the purpose of obtaining nourishment.

It is to be emphasized that the concepts of "producer," "function," and "purpose" have been defined in a manner consistent with the mechanical laws of a natural image. The formal definitions also provide the basis for determining by observation whether or not an individual falls within a given class. It is therefore possible to include questions of function and purpose within methodology, without doing violence to the principles necessary for an answer to questions within mechanics. To the specialist in the field of formal science, there will undoubtedly appear certain aspects of these definitions which will require further elucidation before they can be stated adequately within the language of science. However, within the scope of this essay, this brief analysis of terms will suffice, for the aim here is to show how questions of function and purpose are possible within scientific method.

It is to be noted that the determinacy characteristic of cause-

effect is lacking in producer-product, function, and purpose. In the case of questions concerning the latter concepts, one is really interested in the relationship between two classes of individuals, rather than the individuals themselves. One is interested in the frequency, or "probability," with which individuals of one class produce individuals of another. Questions concerning probability are now definable in terms of the generalization of the natural image to include the producer-product relationship; we can say that the "probability" that x produces y is the limiting value of the relative frequency of all X-elements producing Y-elements, *provided such a limiting value exists within the natural image.* Thus, an individual tosses a coin an indefinite number of times; no single toss is identical with any other, but we can define the limits within which we are willing to specify that the toss was "legitimate" or "honest," and outside of which we reject the toss as "too erratic" to be counted. The (legitimate) tosses therefore form a morphological class, and we are interested in determining the frequency with which they produce a "head." Again, the class of "heads" is not identical in its membership, and we must specify limits within which the state of the coin is really described as a head, and outside of which the state represents a doubtful "fall" (e.g., the coin's coming to rest at a certain angle against something). The class of "heads" is also morphological. We are then interested in the relative frequency with which a toss produces a head. The "true" relative frequency is the probability of the event, provided it exists as a value to be indefinitely approached in accordance with the natural image. (See technical footnote at end of this chapter for a formal definition.)

This simplified description of the nature of probability might seem too much like one we have previously abandoned, for the empiricist also described the probability of an event in terms of relative frequencies. But the difference in points of view is large, is, in fact, all the difference that exists between one who accepts the rationalist's demand for a presupposed theory in all inquiry and one who does not. The "true" probability of an event, like the "true" distance between two points, only has meaning in terms of *both* theory and observation; and just as in the case of the measurement of distance, we demand a natural image to guarantee an indefinite

approach to the ideal or true value, so we demand a natural image for all measurements of probability. It is not, as the empiricist would have it, simple God-given luck that natural events appear to classify themselves into groups for which we can compute a relative frequency with more and more exactness; rather, it is the human-given formalism of nature that demands this sort of regularity for the sake of the meaningfulness of experimental questions. The reason that the relative frequencies must approach some limiting value is that the question of probability is otherwise meaningless; one is "guaranteed" that they do by the natural image which is presupposed in all experimental problems.

Just as in the case of the measurement of the distance between two points, so in the measurement of probability we must be able to make a continuous adjustment of observation to increase indefinitely the precision of our estimates. Let us suppose that we are interested in the probability of death as a result of a certain disease. The question must be made precise by specifying the group of individuals to be studied, and by specifying not only bacteria, but the manner in which and conditions under which they enter the body. These are morphological classifications of the sort mentioned above, and the manner in which they are constructed depends upon the purpose of the question. Now at any given time the number of individuals which can be studied will be finite, and if the meaning of the problem depended upon an examination of these individuals only, then we could not measure the *probability* of death; we could only measure a percentage of deaths within the finite group. As a consequence, the measure of probability demands that the number of observations can be increased indefinitely; as in the case of the measurement of distance, the observations need not be made directly on the object in question. By experimenting on individuals of different morphology, or diseases of a different kind, one can translate the data into information pertinent to the original problem. One of the common methods for performing this translation is to presuppose a certain type of probability law over a whole class of morphologies. For example, suppose the problem is to determine the probability that a certain type of bar-steel will crack under a given impact. If the impact under study is one to be expected in

service, and if the probability of cracking is very low, we may find it impossible from a practical point of view to make any direct observations, for the cost of each test may be high, and to obtain even a rough estimate of a low-probability point demands many tests. Instead, we may conduct our investigations with a very much stronger set of blows; then if we can presuppose a functional relationship between the strength of blow and probability of cracking, we can translate the data into information pertinent to the original problem.[1]

Thus the measurement of probability, like the measurement of mechanical properties, demands a natural image for the continued approximation to the true value. Unless such an image is presupposed, the question about the probability is meaningless, since the answer does not exist. It will be apparent that the image presupposed for probability measurements must be consistent with the image presupposed for mechanical (elemental) measurements; but it should be emphasized that neither image has any priority: if progress, let us say, in the psychological sciences demands abandoning a certain aspect of the natural image, and reconstructing one that conflicts with an accepted mechanical image, the most efficient progress may take place by modifying the latter even though it is acceptable to physicists.

The viewpoint that problems of producer-product and probability are concerned with group properties of the elements of a deterministic natural image enables one to give meaning to certain frequently used concepts of probability-theory. For example, it is evident that "chance fluctuations" or "chance variations" represent fluctuations due to differences of the elements within the morphological classes. They do not represent "freedom" of the elements of nature, but rather the cumulation of effects that are not

[1] This method of adjustment is given the generic name of "sensitivity test"; its use, explicitly or implicitly, is very extensive in experimental measures of probability. The relationship between p, the probability, and the strength of the stimulus a, is usually given by

$$p = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}\, dx$$

where μ and σ are constants that either are estimated or are known a priori. In biological assay, a logarithmic transformation of the stimulus is necessary before this relation can be used. See (1).

pertinent to the problem because our interest is no longer centered upon the individuals themselves, but is centered rather upon ranges of properties of the individuals. Nevertheless, chance fluctuations are not antithetical to the determinacy of the natural image; they represent a state of nature that can be formalized as readily as is mechanics. The error is often committed of regarding problems concerning the nature of chance variations as problems lying outside the domain of exact science; the argument here shows that these problems are solvable within a formal theory of nature.

Again, the concept of randomness can be given meaning in terms of this characterization of experimental method. This concept is peculiar, in that while it plays a role that is basic to most problems of probability, its meaning with respect to experimental method is more often than not left vague or ill-defined. One has legitimate reason to ask for criteria that justify our assuming randomness in Tippett's numbers, for example. This justification is frequently based on the fact that the values of χ^2 do not become significantly high.[1] Such justification might be called *empirical*, and suffers the weakness of all such empirical arguments: since only finite samples are given, we can only at best assert that the χ^2 test does not refute the hypothesis of randomness and not that it "confirms" it. It is certainly not difficult to imagine a set of numbers that will satisfy the χ^2 test, but which are not random; for example, if one selected those numbers he knew would lead to suitable χ^2 values, he would not be drawing "at random."

Another definition of randomness is based on a mathematical construct, and might be called "rationalistic" in that it leaves unanswered the problem of relationships between the construct and a series of observations.

The experimentalist definition of randomness can be illustrated by an example from measurement theory; the experimenter requires that his series of observations oscillate in some random fashion

[1] The χ^2 test, developed by Karl Pearson in a series of classical articles in *Biometrika* and elsewhere, is designed to test (among other things) whether a system of deviations from their respective expected values is unlikely; in the case at hand, supposing the original universe to be known, one can test whether the frequencies given in a table of Tippett's numbers deviate significantly from the original universe from which the numbers were supposedly drawn at random.

about the true value, because if they do not, then the limiting value he would infer from one subsection of his observations would differ from the limiting value he inferred from another. This we take to be the basic idea behind the "Kollectiv" of von Mises (3) and others; each subsection of the observations (selected by a principle that is independent of the observations themselves) should approach the same limiting value. But this requirement is no more than the requirement of control that we have given earlier; we have demanded that at the end of any observation it be possible to determine whether the series is approaching a limiting value, and by this we mean that no subsection of the series approaches a different value from another subsection. The test for control, we said, depends on the presupposed statistical theory; essentially, the formal theory of statistics outlines an image of an indefinite approach to a limiting value, and provides the necessary conditions for such an approach.

We are therefore arguing that the test of randomness of observations is the same as a test of control. In effect, when we *assume* randomness, we assume that our observations are such that a limiting value exists, and the question asked has meaning. Thus, if we assume that a coin is being tossed "at random," we mean that we are presupposing a natural image within which the various elements of the morphological class of "tosses" approach a limiting value in a constant state of control. When we test for randomness, we are determining whether the set of observations is in control and the natural image is permissible.

We have pointed out previously that in questions concerned with relative frequencies of properties in morphological classes, it is necessary to presuppose an image of nature in which limiting values of these frequencies exist. Thus, it may be meaningless to ask for the probability that the face of a die will be six, if the meaning of "toss," the meaning of a "success," and the laws of the natural system preclude any limiting value whatsoever to the relative frequencies; that is, if our morphological classes are so defined that within the laws of the system no limiting value of the relative frequency exists, then it is nonsense to ask for the probability of the event. We would then say that randomness of throws cannot be maintained, meaning thereby that a state of control among the

observations cannot be maintained. Now his situation is perfectly general: when we make observations on a given length, the observations themselves are elements of morphological classes, and the environments in which they are made are not identical but only "similar," i.e., they belong to the same morphology. Our demand is the same: that the relative frequencies of subgroups of the observations approach certain limiting values, and lack of control is inferred if they do not. By assuming the randomness of the observations in any series we are merely assuming that the question asked has meaning: if the observations could not remain in control (i.e., remain random), then the question asked is meaningless.

The result of this characterization of randomness leads to a conclusion that would appear paradoxical indeed to the empirical mind: *the demand for randomness of observation is a demand for regularity in nature.* The empirical notion that randomness is a gift of the uncontrolled nature within which we must conduct our experiments is here replaced by the notion that randomness is a presupposition that must be "put into" nature in order to bring it into control. It has now become a truism of common sense that to conduct a sampling precedure without presupposition is to seek results of a meaningless sort; a public poll of opinion that samples without any prior knowledge of the character of the population will inevitably accumulate wrong results: the limiting values its results approach today may be quite different from those of tomorrow. It has become clearly recognized, therefore, that a randomization demands an a priori assumption concerning the regularity of the universe. It is not always so clearly recognized that we demand more than a principle of selection in these matters: we demand, for example, in the case of polls, criteria of meaningfulness of sentences to various listeners; the poll should certainly *not* phrase the question asked in the same manner for each subgroup if it expects the question to have the same meaning for each subgroup. For example, the phrasing of the question should usually vary with the educational background of the individual.

The viewpoint that a demand for randomness is also a demand for regularity is further illustrated by a second and more specialized use of the concept of randomness in the literature. It is frequently

desirable from the point of view of simplicity so to design an experiment that the observations, or certain sets of observations, are independent of one another. Thus the order in which a series of stimuli are presented to an individual in a psycho-physical measurement should be randomized if possible from trial to trial, lest the fact that one stimulus always precedes another influence the result. To this end, a natural image is assumed in which the desired independence is a consequence (at least to a certain degree of approximation). The assumed image then dictates the ordering of the observations in the sense that it sets down the *necessary* conditions for correct ordering.

The consequence is that one method of randomization used to guarantee independence may not suffice under a change in the nature of the problem (i.e., the classes under consideration), or under a change in the formal presuppositions. There is nothing absolute, for example, about the randomness of Tippett's numbers; their use presupposes that the articles to be tested have been ordered (e.g., by days, or by their position in a box), and that variations are distributed in a certain way over this ordering. If systematic variations occur, the numbers may yield non-random observations. Similarly, "thorough mixing" of granulated chemicals may yield non-random results.

We have outlined the manner in which probability concepts receive meaning under the characterization of experimental method made in the previous chapter. The entire argument might well be summarized by borrowing a phrase from R. A. Fisher to assert our principle, "The more thorough the design of an experiment, the more meaningful is the question asked." We have no doubt put more into the "design" than Fisher discusses, for we have shown that the perfect design is no more than a limiting concept of the true image of nature. Just as the true probability, or the true distance, are limits which control but do not determine observation, so the true state of randomness, or, what is the same, the correct design, are limiting concepts that control our experimental procedure but do not determine it. The perfect design would therefore involve the perfected image of nature; in lieu of this perfection, we make the best approach to it that is possible: we presuppose that

regularity of nature, or, in Fisher's terms, we construct that design, which leads to the most efficient method of testing control.

To summarize: We have attempted to show that probability is a measure of certain properties of an image of nature extended to include morphological classes and their interrelations; that the meaningfulness of questions about probability demands, as do all meaningful questions, the existence of a limiting concept; that the approach to these limits demands a state of randomness, i.e., a natural image from which one may infer the possibility of an indefinite approach to the question asked.

The all-too-brief survey of this chapter does lead to one conclusion which we shall make its closing theme. We started our discussion of experimental method by postulating observation and theory; we end by sketching the manner in which questions about function and purpose, and hence "observation-taking" and "theory-constructing," can be given meaning. We have thus traveled around a large circle, our end-point being the meaning of the postulates we previously assumed. This property of experimental method we take to be inescapable, but the consequences are not disastrous. If an analogy is necessary, all experimental science proceeds in the manner of the physicist who wishes to determine the characteristics of a rather complicated state. He must hold one phase as constant as possible, or, more generally, he must assume the laws governing change in one phase, while he investigates another. So in all experimental work, we assume a construct that enables us to investigate a phase of nature, be it mechanical, physical, or teleological. We may then investigate the soundness of the construct we earlier assumed. We may, for example, take our observations as "given" during one phase of experimental work and not question the assumptions we make about the observer and his "bias," but later we may make these very data the subject of our inquiry. Analogously, we may assume our theory to be sound, and the method of deducing consequences within it, while we are investigating one phase of nature; but later we may investigate the psychological process of theory-construction and the nature of our confidence in a deductive process.

There is nothing "immutable" therefore in the presuppositions we demand for experimental inquiry; there is no true beginning-

point to science. The only invariant to be found is the purpose of all such inquiry; even this is variant with respect to its defining, in the sense that we can become more and more precise as to the meaning of the principle that the end-point of science is the reduction of error to zero.

There is one aspect of the circularity (or, rather, "spiral" form) of science which is to be regarded as the theme of this essay, and which we shall make the closing study in the subsequent chapters. This circularity is based upon a conclusion reached in the study of inference in Chapter II.

The discussion of the preceding chapters has been designed to show that a generalized form of the methodology of Chapter II is the basis for the answering of any question. Hence, we have

1. *The response to any question of fact or law presupposes a method of selection of one among (at least) two alternative hypotheses.*

Since the analysis of Chapter II is general, we have

2. *At least two "distinct" methods of selection exist for any question and any pertinent set of observations.*

In this connection, two methods are said to be "distinct" if the associated risks of error are different (e.g., if for the same type I error, the type II errors differ in some respect).

Finally, we have

3. *A complete analysis of experimental inference presupposes a criterion for the most efficient method of selection;* i.e., *the answering of any question of fact or law presupposes a criterion of value.*

Such a value concept is purposive in connotation, since there are many available and distinct means (the methods of selection), and presumably one final end common to all the methods.

The final chapters of this essay are devoted to the study of the criterion of efficiency thus briefly introduced. Before proceeding to this study, it will be advisable to make several general remarks concerning the experimentalist viewpoint, with respect to certain aspects of science.

Note: The object of this note is to sketch a formal definition of probability within the image discussed in the body of the chapter. The task is to construct a measure of the probability that an element

of a given morphology X (e.g., tosses of a coin) in a specified environment N at a specified time will produce an element of morphology Y in another environment at a later time. We suppose that the elements of the X, N, and Y morphologies can be ordered. If the classes are defined by tolerance limits along a continuous scale, this is a relatively simple matter; e.g., X might be a toss specified by a force lying in the range F_1 to F_2, and by direction of application specified by the range β_1 to β_2. Now for each choice of an element of X and a corresponding choice of a member of N, a complete time-slice of nature (within the closed image) will be determined. Hence, it will be possible to determine by the laws of the natural image whether an element of the Y-morphology occurs in a specified region at a later time. Thus, for every element of the ordered set made up of X and N, there will be paired a 1 (success), or a 0 (failure). The set $X + N$ has a measure greater than zero in the Lebesgue sense (since X and N are defined along continuous intervals). Call this measure m_1. Call the measure of the subset of $X + N$ for which there is a corresponding 1, m_2. Then the probability of a Y-element, given an X- and an N-element, is m_1/m_2.

The sets X, N, and Y need not be morphologically defined; e.g., X might be a functionally defined set. But for the definition to apply, the elements of the set must be capable of being ordered. If X and N are mechanically defined, then there is only one possibility, and the true probability is taken to be either 1 or 0. Evidently the measure of probability depends on the specifications of X, N, and Y. The probability that I shall die at age sixty, when I am considered as a member of the set comprising all citizens of the United States, is not the same as the probability that I shall die at age sixty, when I am considered as a member of the academic profession.

The measure of probability defines the "true" probability. The procedure of estimating the true probability, and indefinitely approximating its value, is the same as that already discussed for true lengths. Thus, as the sample size increases, the relative frequency, or an adjustment of the relative frequency, of successes to total trials, should approach (stochastically) the true value. Where this fails to occur, the natural image and/or the observations must be revised.

REFERENCES

1. Churchman, C. W., and Epstein, B, "Tests of Increased Severity," *Journal of the American Statistical Association* (1947).
2. Cowan, T. A., "Towards an Experimental Definition of Criminal Mind," in *Essays in Honor of E. A. Singer, Jr.*, Univ. of Pennsylvania Press (1942).
3. Mises, R. von, *Probability, Statistics, and Truth*, Macmillan (1939).
4. Singer, E. A., Jr., "Beyond Mechanism and Vitalism," *Philosophy of Science* (1934).
5. Singer, E. A., Jr., "Logico-Historical Study of Mechanism," *Studies in the History of Science*, Univ. of Pennsylvania Bicentennial Conference (1941).
6. Singer, E. A., Jr., "Mechanism, Vitalism, Naturalism," *Philosophy of Science* (1946).
7. Singer, E. A., Jr., *Mind as Behavior*, Adams (1924).
8. Singer, E. A., Jr., "On Spontaneity," *Journal of Philosophy* (1925).
9. Singer, E. A., Jr., "On the Conscious Mind," *Journal of Philosophy* (1929).

Chapter XIII Applications of Experimentalism

The survey of the nature of experimental method that has been made in the preceding chapters must have left many questions unanswered, as all formal schemes do. And it would not be unnatural to ask whether the experimentalist viewpoint represents any significant progress in man's attitude towards certain problems, and in general it would be obvious enough to inquire whether what has been said in any way affects the method of inquiry in science.

We have attempted, of course, to represent the progress attributable to experimentalism by formal arguments; we have exhibited a classification of opinions as to the nature of experimental method, have claimed that these opinions have more or less followed a certain sequence in history, and have tried to show the formal meaning of progress in this sequence. Thus experimentalism is made to represent a stage of synthesis of the reflective mind's demand for "certainty" and its demand for "method," the formal pattern attempting to define what these antithetical poles mean.

In this chapter we shall discuss the results of the formal viewpoint with respect to certain problems, in the hope that thereby a clearer understanding will be gained of the insistence on a new viewpoint in the philosophy of science.

1. *Meaning.* As a first example, take the problem of "meaning," which has well represented the antithesis of the two types of school, the certainty-seekers and the methodologists. To the nonexperimental philosophers it seems legitimate to inquire into the nature of the "real," the "good," the "mind," independent of experiential

criteria, for more general criteria than experience seem to exist, namely, those establishing certain and immutable principles. It appears to a rationalist temperament that one may inquire whether a mind exists independent of experience, whether reality is a mental product entirely, whether "being" is a relational concept, etc., provided these questions receive meaning in terms of some well-established fundamental principles.

To the empirical temperament, however, the purely formal can never be said to have meaning; for him, the question asked must be translatable into a definite experience or set of experiences before it can be said to be meaningful. This is the fundamental viewpoint (though variant with the meaning of experience) of the classical positivism of Comte, of the modern logical positivists, and of the operationalists. For example, one inquires whether the ether exists; such inquiry is meaningless unless the question can be so translated that a test can be put to experience. The idealist versus realist argument is another case in point. Can we assert the reality of a world independent of any observer? Rather than pose an answer, says the positivist, first put the question to the test of meaningfulness. Is there any manner in which such a question can be verified in experience? If so, the answer is to be found there; if not, then the question is meaningless, and so must be all disputes concerned with it. Again, one might inquire with Shewhart (9), what one means by the question "Has this sample been drawn *at random* from its population?" It does not suffice to give a purely formal explanation of randomness. If purely formal explanations are the only ones possible, the concept is really meaningless to the experimenter. He cannot decide on purely formal grounds, in any practical instance, whether his observations are random. Randomness, for the empiricist, must be translatable into definite physical operations, or else it must be regarded as a meaningless concept.

In general, the mind that inquires into nature, and finds such inquiry the only meaningful kind, seems forced to accept an experiential criterion of meaningfulness, and to regard many of the classical problems of science, of art, of ethics, and of religion, as meaningless, since apparently no success could ever be made in showing how such questions can be put to the test of experience.

Few can doubt the healthy impact that the positivist position has had upon modes of inquiry; it has sharply distinguished the schools of thought, and has raised a standard under which the proponents of experimental method can fight their battle against a reactionary movement. To return to a pre-positivistic viewpoint is to return to a pre-scientific viewpoint, to become as reactionary as an advocate of the indisputable power of the sovereign in the eyes of one with a democratic outlook.

But now is the time to inquire whether positivism has outrun its usefulness, and whether, in order to save the terrain that has been gained under its influence, we are not forced to generalize its position. If positivism can no longer be adequately defended, then it becomes imperative to restate its position lest its weakness lead to its complete overthrow in favor of the reactionary movement.

The weakness of positivism lies not so much in its wholesale dismissal of problems that throughout the ages have worried many great minds. Such dismissal, despite its implied devaluation of much of present-day philosophical thinking, merely emphasizes the antithesis between the schools. Nor does its weakness lie so much in the obvious formal argument that can be turned against its fundamental tenet by inquiring whether this tenet has positivistic meaning (i.e., can the question whether meaning lies in a definite experience be put to the test of experience?). Such logical refutations rarely sway either side of a dispute.

We have already mentioned one of positivism's fundamental weaknesses: an apparently futile attempt to reduce the generality of definition to the specificity of sensation; thus, Shewhart at best gives *one* operation that defines randomness, and admits it to be only one. The definition of randomness is therefore not any one of these examples, and its true meaning must lie in a nonoperational concept. In terms of previous discussions, "operations" are only means to an end, and no single means is *necessary* for the pursuit of an end.

But perhaps the most significant failure of positivism lies in its attitude towards a problem that the advance of experimental method has made more and more pertinent. As the science of psychology has improved its experimental techniques and the formal theory behind these techniques, it has become increasingly clear that the

term "experience" or "observation" is a highly complicated one, at least as complicated as the physical concepts of measurement, or length, or mass. It turns out that a simple sensation is usually an ambiguously defined set of impulses and reactions, towards the clarification of which we can as yet only make a beginning. If this be the case, and few acquainted with experimental psychology can deny it to be so, then it must appear absurd to reduce the meaning of the concepts of the other sciences to certain types of observation or sensation. This is simply a case of robbing Peter to pay Paul. We naively assume that one and all have a good basic understanding of the meaning of observation, or experience, and proceed to define our concepts in terms of these bases; we could as well reverse the procedure and assume that everyone has a sound notion of the meaning of "random," or "mass," or "space," and require a reduction to these concepts in the proper definition of a term. And such vague generalizations concerning the meaning of observation as one finds for example in Lenzen's *Procedures of Empirical Science* (3) miss the important point; they suppose that the difficulties inherent in an exact analysis of the meaning of observation can be removed for the most part by verbal references, as though, à la Socratic dialectic, everyone really knows what an observation is, and a little or a lot of question, answer, and illustration will remove the trouble.

The problem of observation is far more difficult than this; its solution will require an advance in psychology of as revolutionary a character as the Copernican advance was for astronomy. One might safely assert that the reason so many people feel certain about the meaning of observation is that they know so little about it, just as in the pre-scientific ages so many knew what "mass" and "space" meant, simply because they knew so little about such concepts.

The point of view we have tried to develop in this essay is that the time has come to recognize the circularity, or spiral form, of science, and the complete interdependence of the sciences. It is perfectly proper to consider one phase of nature as though it were known, while we develop another phase, as long as we do not make this a permanent state of affairs. Operationalism, then, is a convenient method for removing certain ambiguities, but it can in no sense be regarded as the final answer to the problem of experimental method.

There must come a time when we investigate experimentally the meaning and properties of direct experience, physical operation, observation, sensation, and such.

Instead of a reference to experience as a criterion of meaning and truth, we have suggested the criterion of a measurable error, the magnitude of which can be reduced indefinitely to zero. Thus, a question of fact is said to have meaning if (1) we can frame a response, (2) measure the error of the response, and (3) reduce the error within any given distance of zero. This is not to assert that the criterion we have proposed is any more "obvious" as to its meaning than the criterion of positivism. Its advantage over the positivist formulation is that it does not make any one science basic to all experimental method. One may raise the question what "response," "error," "reduction," mean, and may presumably give responses to such questions and measure their errors. No requirement of fixed reference points is demanded, and yet all the advantages of a positivist insistence on experiment are retained. Further, we are not forced to discard as meaningless whole sets of questions, simply because an experiential reference point cannot be found. The true nature of reality can become a meaningful problem for discussion, despite the fact that reality is never directly observed; for we may define the "real" world as a limiting concept, toward which all experimental effort is proceeding. The "true" and the "real" represent the ideal (never attained) answers to all questions; they represent the state where error has been reduced to zero. Similar translations of many metaphysical and epistemological problems lead us to conclude that these fields of inquiry, far from involving meaningless problems, may be regarded as phases of science (say, psychology) provided their problems can be formulated according to an experimental criterion of meaning.

2. *Tests of Hypotheses*. The experimentalist viewpoint can be made to shed a new philosophical light upon the meaning of statistical tests of hypotheses. Experimentalism insists on including among the assertions of the alternative hypotheses *all* the presuppositions made by the experimenter (i.e., all he can make explicit). The formulation of Chapter II merely demands that the hypotheses be of a statistical character; no consideration is given to the problem of the

source of these hypotheses, since the mathematical statistician is chiefly interested in how to proceed once the hypotheses are formulated.

The experimentalist viewpoint is that the statistical hypotheses are (or should be) consequences of a formal theory of the science. They are theorems of contrary formal systems of the science, deduced according to the criteria of formal method. Only when the hypotheses are so considered does an experiment have meaning; the error, so commonly made, of trying to put the entire burden of meaning upon the statistical hypotheses always leads to failure. One cannot simply take a set of data, make certain distribution hypotheses about their populations, and proceed to a statistical test; one cannot do so and expect that a meaningful answer will be the result. To paraphrase Kant, statistical tests without theory are blind: no general results can be asserted, no predictions made, unless one assumes that the statistical hypotheses are consequences of a general theory within which prediction can be made independent of specialized restrictions. The analysis of experimentalism also places certain demands upon this general theory: it must guarantee the existence of an answer to the question asked.

We may therefore take the following to be the criterion for the meaningfulness of statistical tests: *every statistical hypothesis should be a consequence of a formal theory of nature.* This principle means that hypotheses concerning normality of distribution, randomness of observation, difference in means, etc., should all be theorems of some theory of the science in question; only in those cases where we can explain the meaning of the statistical hypotheses in terms of some general considerations can the statistical test be said to have meaning. This, for example, we take to be the real intention of the operationalist analysis of statistical tests: operationalism insists that the concepts of randomness, probability, etc., be explained in terms of certain physical operations. The emphasis here should not be upon the certainty of the physical operation, since such certainty does not ever exist; the emphasis should rather be upon the necessity of tying down the meaning of the assumption of randomness to certain assumptions regarding the behavior of objects in the physical world. When we say that the assumption of randomness is equiva-

lent to drawing chips from a bowl in a certain manner, we are attempting to characterize in a formal manner the way in which the objects (coming off a machine, say) are produced in time, or, more precisely, we are attempting to characterize formally the type of sequence of measurements to be expected as the objects are measured in a certain order. Thus, as we have stated earlier, the assumption of randomness is an assumption that expresses a certain behavior of objects in the natural world; it is a sufficient condition for experimental control, and may vary considerably depending on the science and its formulation.

The experimental situation is then described *ideally* as follows: a formal theory is developed within a given science in the usual fashion, as described above in the chapter on rationalism. The theories all contain the concepts of observation, and such statistical terms as distribution function, randomness, probability, and so on; these statistical concepts have the same formal meaning within each theory. The theories are also so constructed that they have a certain set of assumptions (and consequences) in common. These assumptions we call the "invariant assumptions," or, in our previous terminology, the *presuppositions* of the experiment. The invariant assumptions may contain distribution assumptions about the observed values of certain objects at certain times, about their randomness, together with assumptions about the accuracy of measuring devices, etc. Then, when a question has been proposed, we so translate it that we can deduce within each theory a certain statistical hypothesis concerning the observations at any given time. The resulting alternative hypotheses are "tested," by deducing from each a judgment concerning the probability of a specified event, and the choice among the hypotheses depends in part upon statistical theory. We decide to accept one of the original theories of the science, not on the grounds of consistency, but (in part) on the grounds of probability.[1]

This marks the chief distinction between what we have called the formal and nonformal aspects of science: the criterion of acceptance or rejection in formal method is based on consistency, whereas in nonformal method the criterion is partially based on probability.

[1] The criterion of choice also depends on loss and risk; see Chapter XV.

This enables formal theory to restrict itself to fields where probability concepts are not involved, but this does not imply an absolute independence of the formal and nonformal. The distinction between the two aspects of science is a relative one, for, as we have shown, it is meaningful to investigate nonformally the problem of consistency, or, in general, the problem of deduction in formal science. In such cases, the investigator presupposes a certain psychological or sociological theory concerning the purposes of the mathematician, and proceeds to investigate them. The "circularity" involved need not cause a difficulty, as the discussion in Chapter XI has shown.

The above, be it noted, represents an *ideal* account of the experimental procedures; it is what should be done if the "best" method is to be followed. The account is simply another way of stating the position expounded in Chapter XI. The emphasis here is on the statistical aspects of experimental method. The account given previously attempts to define the general question that science asks, the question, indeed, that defines scientific inquiry: does experimental control exist, i.e., are we approaching the true value in a satisfactory manner? The present analysis attempts to define the best procedure for answering such questions.

It will be noted that the "data" or "given" of an experiment are simply the invariant assumptions we have discussed above; they may be particular (individual) or general, and need not be the same for all investigation. The error of the "conservative" schools of rationalism, empiricism, and criticism lay in their attempt to find one set of invariants for all inquiry. The lesson of relativism is that no such set can ever be found, for one may always investigate at a later time the assumptions one makes now; that is, one may take an assumption that is the same for all the alternative theories in this experiment, and later on construct a set of theories in which the assumption is denied in various ways. "Inquiry," then, refers to the properties of the alternative theories that are not the same, "presupposition" to the properties that are the same; inquiry and presupposition are the correlatives of experimental method.

The comparison of formal and nonformal science may be made as follows: in formal science the undefined concepts are not restricted, may be translated into those of another science, and need not entail

any reference to particulars; the criteria of consistency and rigor of deduction represent the truth criteria. In nonformal science, alternative formalizations of the science must be made which differ in at least some of their properties; these formalizations must be rich enough to include statistical concepts concerning collections of individuals, and hence rich enough to include individuated objects. Each formal theory must thus include assumptions about a series of observations, some of these assumptions being common to all the theories, others not. The criterion of selection among the alternative formal theories depends on probability judgments, and, more generally, upon the criteria of "efficiency."

The thesis developed here on the interrelationship of formal and nonformal science undoubtedly demands elucidation. Certainly the experimenter is not aware in general that his observations are "postulates" assumed for the sake of proceeding to a response; certainly, too, the statistician is often unaware that his hypotheses have any definite relationship to the general theories of the science within which the experimental problem arose.

Suppose, as an example, that the experimenter asks whether one type of steel is harder than another. The operationalist has made clear enough that, in order to give an answer to the question asked, we must specify the meaning of "hardness" in such a manner that the observer can perform a definite set of operations. These operations, however, do not occur to the experimenter in a haphazard manner. The original question asked has some definite meaning to the asker, who presumably will take some sort of action as a result of the experiment. He may desire a steel which will not scratch or dent easily; or he may desire one that can stand up under extremely heavy loads; or one that can stand sudden shock. The operations we perform to determine which steel is the harder will depend upon the purpose of the question, i.e., upon the action we intend to take as a result. A Rockwell hardness value may be the pertinent datum, or it may not; if it is surface hardness that is of interest, the Rockwell will serve, but a Rockwell reading might give no information whatsoever if what matters is the ability to stand shock. In the light of this illustration, few could assert that it is simple and obvious to decide what observations are pertinent to a problem. To find out

whether a datum of sensation is pertinent to the problem requires the background of an expert, and even the expert is never absolutely sure that a given method of collecting data is really pertinent to the answering of a question.

But suppose our metallurgist could be sure that a Rockwell reading was a pertinent observation, and hence had confidence that a certain operation, if carried through correctly, would provide information concerning the question asked; there still remains the problem of whether a given reading was made correctly: whether the operation was performed exactly according to directions. Anyone well acquainted with the problem of making a series of similar observations will have no difficulty in recognizing the point here; even the "simple" operation of deciding whether an explosive fired or not, when repeated many times, becomes a difficult one to control in the sense that all explosions and nonexplosions are counted correctly.

This point has been illustrated over and over in so-called "round-robin" tests, which are designed to determine whether a group of laboratories, which are supposedly conducting the same types of measurement, come out with the same results. Very specific instructions are drawn up which each laboratory is supposed to follow. And almost invariably, the laboratories show a greater difference than can be accounted for by the errors of the measurements. The usual procedure is to try to patch the matter up; but there is apparently some real basis for differences, possibly arising out of the training each laboratory gives. One could expect the physicists to try to ignore this problem, since it takes them out of their field. But it is very surprising to hear modern psychologists refer to "simple" operations, or operations based on a "common agreement."

Thus, as a result of this discussion, we have

a) all questions that have meaning are asked for the sake of some purpose;

b) it is impossible to decide without error whether or not a given operation is a means to the purpose for which the question was asked;

c) even though a type of operation does serve to provide pertinent information on the question, it is impossible to decide without error

whether any particular observation belongs to a specified class of operations.

Observation, therefore, must become as much a postulate of our method as the general laws of the system; a later science may come to decide that observations made today are not pertinent in the way we took them to be, just as a later science may come to decide that a theory of today is not pertinent to the problem, or not pertinent in the way we took it to be.

Suppose, now, that our experimenter has taken a series of Rockwell readings on steel bars of each type. What then? How can these observations be made to serve as means to the end of responding to the question? At this point the statistician is called in, and is handed the data with the request for an "analysis of the results." What is often done is to turn to a certain page of a statistical text, and apply the formulae listed there without further inquiry into the original problem or the method of collecting the data. This procedure is the basis for "figures don't lie, but statisticians do." If statistical procedures are used only in analyzing a set of data, then we will get out of these procedures just what we put in; we will show that if a set of numbers are given, and if we assume them to be a certain type of sample drawn from a certain kind of universe, then certain properties can be ascribed to them. But the relationship of all this to the original question will be unknown, and the statistician's work has become a mere arithmetical exercise. It cannot be too often repeated that the statistical procedures of inference must be dependent upon the nature of the question asked, and hence upon the formal theory the experimenter brings to his problem.

The consequences of overlooking this dependence of statistics upon prior knowledge has been the basis of a strong distrust on the part of experimenters of the exact statistical methods. Simply to analyze a set of data without knowing the conditions of the experiment and the meaning of the problem is an exceedingly dangerous procedure; we are apt to translate an answer that has only statistical meaning into an answer that forms the basis of a certain action (e.g., using one steel rather than another), when the translation is wholly unwarranted. In order that statistical procedures be experimentally sound, it is necessary to postulate that the statistician's hypotheses

are "pertinent"; that is, we must know why randomness can be assumed, or why a continuous distribution function can be posited. And the answers to these questions lie in the meaning of the original question and the techniques for gathering data; but this meaning and these techniques must be given within a theory of the science in terms of which the original question is posed. Hence, statistical hypotheses should be consequences of some such theory of nature. The result is that a close cooperation is demanded between the statistician and laboratory worker; actually, the most successful applications of statistical techniques occur when the statistician is a member of the research group, and follows the problem through from the beginning.

3. *The Unknown.* The analysis we have given of experimental method will not entirely satisfy the reflective mind, for it does not make direct reference to the unknown factors that exist in any experiment, to the intuitions and guesses of the experimenter that appear to fill the gap where an exact method is lacking.

It is the necessity of lack of method in many cases that has frequently suggested a more general criterion of truth than we have offered. Thus, within probability theory, the question has arisen as to whether "relative frequency" can be made the sole basis of our confidence in a given theory. The analysis above shows the manner in which probability is made to play a deciding role in experimental science: we deduce within each theory a theorem stating the probability that (i.e., the long-run frequency with which) a given function of the observations will occur. If this probability is very low, we abandon the theory. It is argued that in many instances we are not converted to belief by a method of this sort, but by more general criteria of which this method is a special case. I do not investigate the dangers of crossing a crowded street against the light, or in forgetting my rubbers on a rainy day, or in excessive drinking, by considerations of the sort we have been describing, and yet I do appear to formulate probability judgments about the consequences of these activities. If so, then there must be a more general meaning of probability, and a more general truth-methodology.

The dispute certainly appears to be a verbal one, and it is somewhat surprising that it has taken up so many pages of the philo-

sophical journals. Within the science of psychology there are evidently certain problems connected with belief and the environmental conditions that induce belief. These may all be investigated according to the criteria of investigation we have given above.[1] But that the psychological concept of belief should be made the basis of truth simply because a certain form of belief is necessary in every investigation is as absurd as making the concepts of logic the basis of all meaning because these concepts are necessary in every definition. The dispute simply seems to be concerned over what problems we are to call basic to all experimental investigations (the philosophical problems) and what problems are problems of the science of psychology.

In general, it always becomes necessary for the experimenter to investigate the procedures he has used in his work; in such cases, he may find that a step which appeared "obvious" before is actually in error. "Intuition" in such cases is raised to a conscious or methodological level; the intuitive aspects of any investigation belongs to the invariant assumptions, and in so far as they remain unexpressed, they are assumptions that are not even potentially investigated. All the expressed presuppositions are those that later on we may intend to investigate, i.e., to change among the alternative theories. Hence, *an assumption is made intuitively if no intention to investigate it exists: the intuitive is the potentially changeless.*

4. *The Laboratory.* The laboratory has traditionally been taken as the aspect of science which represents control. The purpose of a laboratory is to design conditions under which the collection of data is unmolested by unknown influences of any large magnitude. The predominance of the laboratory in techniques of control has led to the erroneous conclusion that the laboratory is a *necessary* condition for controlled experiment. Hence, the tendency has been to regard sociology, and to some extent psychology, as fields incapable of becoming as experimentally exact as physics, since within sociology it will be impossible to set up adequate laboratory controls.

The sanctity of the laboratory has thus led to an extremely restricted view of experimental science. It has led the physicist to regard his field of investigation as a model; only to the extent that

[1] See Chapters XI and XII.

other fields imitate the model can they be said to be experimentally sound. The biologist should apparently attempt to reduce his laws to differential equations, susceptible to experimental (laboratory) check; where he can, the sociologist should apparently attempt to set up "laboratory groups" for study, eliminating as much as possible outside influences that might affect the behavior of the individuals under study.

Thus, the physicist supposedly has nothing to learn from the sociologist concerning experimental control, and the sciences become uni-directional in their hierarchy, depending on their degree of "precision," i.e., upon their similarity to mechanics and physics. Such a viewpoint has led most contemporary philosophers of science to look for the meaning of science within the realm of physics, and to moralize in general from what holds within the physical sciences in particular.

All this restricted analysis is based upon the assumption that the laboratory provides the necessary conditions for adequate experimental control. With the reawakening of interest in methods of control that has occurred in the past twenty years, especially in industry, we have gradually been led to abandon the older concept of a laboratory science, and to redefine the meaning of experimental control in a much more general manner. We have been led to this revision on two counts: (1) it is no longer obvious that laboratory techniques are *sufficient* for answering critical questions, even in the physical sciences, and (2) it is no longer obvious that the laboratory is *necessary* for obtaining a set of reliable data on a given question.

As an example of the first of these two doubts of the complete adequacy of the laboratory, suppose an industrial research worker asks whether a certain type of steel will be strong enough to stand up under usage in an automobile. The "standard" procedure would be to devise laboratory physical and chemical tests which would presumably be designed to provide the required assurance. In recent years the standard laboratory method has come in for considerable attack, for it is argued the laboratory tests are conducted under such highly specialized and arbitrary conditions that the predictions we make from the data are never reliable unless "confirmed" by actual practice. In other words, the attempt to control

which characterizes laboratory procedures, gives us such restricted information that the data cannot really be said to be answering the question asked. In general, it looks as though a reply to a question (or a check of a theory) that is conducted within the narrow domain of the laboratory provides very incomplete information about the natural world. The opponents of the laboratory method of test insist that reliable information must be obtained under the conditions of actual service, or, in the case of the more theoretical sciences, under the most general conditions in which we expect the theory we are checking to hold.

To this attack on the laboratory by the practical-minded empiricist, the technologist replies that the "conditions of actual service" are never precisely enough defined to provide us with any reliable data. It may be true that the laboratory tests on a certain type of rubber are done under specialized conditions, but what good does it do us to run the tire around a gravel track until it wears down? Does not the amount of this "service" wear depend upon who drives the car, how he turns the corners, what type of gravel we use, what the weather is, and so on? Hence, the service conditions are as highly specialized as the laboratory, and the only difference is that we do not know all the conditions in service tests.

The work that has been done in experimental design in statistical theory, together with more advanced research in methodology, has shown how this conflict between the laboratory test and the "actual conditions" test can be resolved. The statistical contribution has consisted in showing how an experimental problem can be handled when several variables are changing, so that it is no longer necessary to keep all conditions constant except the one under study. Formerly, if one were investigating the effect of moisture on the fertility of a certain type of seed, for example, he would have tried to use soil and temperature conditions as nearly as possible alike, while he observed the seed fertility with varying amounts of water. Such a laboratory test, however, might be quite worthless to the farmer, for the atmospheric conditions under which it was conducted might very well restrict the application of the information contained in the data. Thus the "fertility versus moisture curve" we obtain with one temperature may change enormously when we

choose much colder or much hotter climates. The newer techniques of experimental design, analysis of variance, analysis of covariance, and correlation analysis enable us to vary as many of the factors as we consider important and still derive exact inferences applicable to "actual situations."

The newer statistical techniques, however, do not set down the general criteria which specify *what* variables are to be changed; nor do they specify the operational criteria guaranteeing that those aspects of nature which we are not actively controlling have no pertinent influence on the results. These criteria are supplied by fulfilling the general methodological demand that the experimental problem be as completely formalized as possible; specifically, the demand is for a complete set of conditions under which a set of operations may be said to provide adequate answers to a problem. This means that the experimenter must become precise enough as to the meaning of his question so that a set of specific observations can be judged regarding their adequacy, and become general enough in his presuppositions about the natural order so that he can decide whether a continued sequence of observations would approach some limiting value. If these formal conditions are satisfied, we can say that the experiment was "controlled" in a formal sense. The demand for such formalism on the part of the experimenter is a "big order," to be sure; in many cases we are not prepared to set down all our formal presuppositions, and it would be foolish to expect that this demand can actually be satisfied in the early stages of scientific developments. But the process of discovery becomes scientific if and only if such control exists, regardless of whether the investigation takes place inside a laboratory or not.

On this account, there is no reason at all why investigations within the social sciences should not be "controlled," despite the impossibility of a social "laboratory." The social scientist, let us say, wishes to determine the preferences of a certain social group with respect to a government policy. He may decide that pertinent observations can be collected by a questionnaire delivered to a sample of the population. If so, he must specify formally the conditions under which a sample can be called "adequate"; these conditions in effect will characterize the question he is asking.

Must the sampling be done within a certain time interval? If so, then his question has a certain temporal significance; if not, then the time variable must be assumed to have certain characteristics. Must the sample be selected by a random number scheme over certain directories? If so, then the question asked means that we are to measure the opinion of a population defined by such directories; if not, then some alternative sampling scheme will be used which will lend a different meaning to the question.[1] Further, the wording and presentation of the questionnaire must be assumed to provide pertinent data. This means that formal presuppositions must be made as to the meaning of an individual preference, and the conditions under which it is measurable. Does verbal acquiescence to a question imply an "actual" preference? If so, then preference is an oral characteristic, and a knowledge of preference implies a knowledge of how a person will express himself, and not necessarily a knowledge of how he will otherwise behave. If not, then other criteria of preference must be stated which will lend formal meaning to the term. Questionnaire construction, if done *scientifically*, presupposes a formal characterization of individuals and their responses. The mistake of so much of present-day psychometrics lies in ignoring the problem of the precise presuppositions of questionnaire construction. Many of the so-called "discoveries" are simply a result of implicit assumptions the experimenter "put into" nature in framing his questionnaires.

As to what choices we must make in selecting the presuppositions of controlled inquiry, these depend upon our initial purposes. Such purposes can certainly be the object of another inquiry, and usually will be, if the experiment is to have any value. The circularity that results need not trouble us, at least from a logical point of view, if we maintain the experimentalist postulates of method. For we may have an inquiry A that presupposes for control purposes a certain result of inquiry B, and later we may reverse the procedure, without violating the basic principle of a continued progress towards a final end.

The laboratory-minded scientist may wish to have one final word. He may argue that the presuppositions we use in the laboratory can

[1] For a further discussion of this example, see (4).

be assumed with a great deal more confidence than the presuppositions we make in the study of social attitudes. Ought we not to add the condition that true control only exists when we make presuppositions about which we are "reasonably" confident? The answer is that the proper selection of presuppositions is one of degree. If scientific organizations were rich enough, they could run concurrent inquiries on our presuppositions, and the presuppositions of presuppositions. The fact that we cannot do so does not invalidate the method, as long as we keep the nature of the presuppositions explicit.

But the real point of the laboratory-scientist's objection, if generalized, represents a significant problem of experimental method. We have said that, in order to answer any question of fact, we must set down the formal conditions which describe an observation and its pertinence. But what are the criteria which govern our choice of presuppositions? Under what conditions shall we presuppose randomness of observation, or presuppose that the reaction time of an observer is satisfactory, or that certain types of observation are unreliable?

This is a question of our choice of "law" in the natural order. In order to conduct controlled experiments, we must make certain provisional answers to general questions about natural events. Any such general question will be represented by a set of alternative hypotheses, one of which we must select in order to proceed.

Now this question of the choice of presupposition is itself a specialized form of the following problem. Granting that science never attains its ideal of absolute confidence, then how should we act on the basis of our present observations with respect to the urgent problems of our day? If science is to be pragmatic, it must serve as an instrument for our needs; but if science's instruments can only be used when science is satisfied with their perfection, then they will never be used. Since we can never *know* a fact or a law, how can we take action on the basis of science? And if science is not the basis of our actions, then does not the experimentalist viewpoint depreciate the value of scientific behavior? These are problems concerned with efficiency which we have reserved for subsequent chapters.

5. *Indeterminacy.* The discussion of the meaning of "answer" which is contained in the previous chapters may cause some to wonder whether in this sense science can ever provide answers to certain questions, in view of the results now commonly accepted within quantum theory, and expressed by the Heisenberg "Indeterminacy Principle." Much has been written on the quantum theoretic principle of indeterminism, and it would take us far afield to review all the arguments waged over its "truth" and its significance. Our present discussion will be a rather modest one in its aim; we shall simply say that *if* the present formulation of the indeterminacy principle implies the nonexistence of an answer to certain questions in the sense defined in Chapter X, then either the present formulation of the question must be abandoned, or else the question must be meaningless within physical theory. This is a rather obvious consequence of the theory of meaning developed here; it is a rather modest claim in the sense that the indeterminacy principle does not represent a "critical" case against the present theory of meaning. In the first place, we do not think it obvious that the principle, as it is now understood within physics, implies the nonexistence of answers. We do feel, however, that entirely too little effort is being expended in the study of the meaning of the principle. This lack of logical analysis arises, I think, out of the fact that the untenable consequences of a common-sense interpretation of the principle are not apparent to the scientists themselves. If we assert that, as the error of measurement of position decreases, the error of measurement of momentum increases, then as methodologists we have the right to ask what it means to assert that at a given moment of time a particle has *both* position and momentum. We can certainly construct a formal image of nature within which it is meaningful to talk about both the position and momentum of a particle; but the analysis of the history of methodology has shown that *purely* formal defining has no meaning; that the validity of formal definitions is itself a matter to be investigated experimentally. Consequently, since we cannot talk meaningfully in general about both position and momentum, under one formulation of the indeterminacy principle, it follows that it is also meaningless to talk about our "inability" to measure both.

This has driven some who have taken this consequence seriously into very strange ways. Thus Reichenbach (7) has felt the result to be strong enough to force a revision of fundamental logical categories. It may be that a crisis will sometime arise within science which will force us into such drastic measures, but we should certainly attempt other more amenable remedies before we start rebuilding the entire foundation of method. The introduction of a third truth-value into science would affect in a very drastic manner the whole of present methodology; for example, it would vastly complicate the theory of tests of hypotheses presented in Chapter II. One feels that Reichenbach's remedy is analogous to strengthening certain of the foundation pillars of a building by taking them out of their proper places and sending them off to the shop for remodeling. Before beginning to remodel science along indeterministic lines, it would be better to study first (at least) the following possibilities:

1. Does the indeterminacy principle represent an inadequacy of *present* methodology within physics? If so, i.e., if the principle is stated to show that within the best developed techniques of present experimental physics, we must temporarily forsake giving more and more precise responses to certain questions, then the principle, far from being "revolutionary" in character, and far from forcing the methodologist into untried ways, is actually only one instance among thousands of similar experimental predicaments. There are examples without number in the chemical, biological, psychological, and sociological sciences where at present we find it impossible on theoretical grounds to separate out the errors due to the techniques of measurement, and where an increase in the number of observations always entails a positive bias, which cannot be estimated. An example of this sort of thing, which we will want to discuss more fully later, is the measurement of the productive efficiency of a social group, where the measurements must be conducted by the group itself. We do not mean that this example has the *same* formal character as the measurements on particles; but the analogy is close enough for us to ask whether there is anything very "new" about the indeterminacy principle, whether, in fact, the principle does not simply represent a stage in the process of measurement which seems to be more or less inevitable, and which a future science

inevitably overcomes. Actually, to show that this viewpoint on indeterminacy is "wrong," it would be necessary to accomplish the seemingly impossible task of "proving" without risk of error that *no* future developments within experimental physics can overcome the implications of the indeterminacy principle (8).

2. The second suggestion with regard to indeterminacy is concerned with whether or not the "newer" physics will wish to maintain the same basic dimensions as the classical in its description of a physical state. It is true that a certain convenience of language results by continuing to talk about the "positions" and "momenta" of "particles," but this nomenclature may also be misleading. It certainly does not seem unreasonable to assert that a formal revision of the criteria of "physical state" will lead to the maintenance of complete determinism, even within the present quantum theory (5).

3. Finally, it might be pointed out in passing that modern statistical theory has developed rather amazing techniques for allocating "sources of error," provided certain presuppositions are made concerning the distribution functions of the observations. These techniques might be applicable to the problem of measurement within atomic physics, so that the "error due to uncontrolled causes" and the "error due to the techniques of observation" could be separated.

6. *Cooperative Research.* As a concluding reflection on the analysis of experiment, it is important to note the kind of unity of science that the analysis implies. In that all expressed phases of an experiment may be investigated, a complete interdependence of the sciences is a consequence. The physicist must eventually call upon the psychologist to aid in the analysis of the errors of observation, and upon the sociologist to determine the best conditions under which experiments can be made. That the sociologist and psychologist have to rely upon all the other sciences in the formulation and investigation of their problems is well known. According to empiricism or rationalism, it might indeed be possible to set up a permanent hierarchy of the sciences, but after relativism it must become apparent that all such hierarchies are relative.

As the progress of science brings a realization of this interdependence of the sciences, the problem of cooperative research becomes more and more pressing. No one can expect to be an expert

in every field: e.g., few physicists can ever hope to know enough of mathematical statistics to avoid all the pitfalls of application; few chemists can hope to know enough physical theory to account completely for the mechanism of chemical change; few scientists can hope to know enough formal theory to make sure of the accuracy of deduction. If this be so, then the only solution is that the scientific mind become a group mind; the unification of science is necessary for the progress of science (1) and (2).

With this in mind, namely, that the research worker in any one field is going to be constantly in need of the direct cooperation of workers in other fields, it is highly important that universities begin to recognize the necessity of cooperative groups. The object of such groups should be: (a) to develop exact methods of experimental design that will be simpler to apply, will increase the precision of results obtained from a given set of observations, and will reduce where possible the number of observations necessary to give a definite response to a given question; (b) to coordinate the research in (a) with the most pressing present-day problems of the various sciences; and (c) to apply whenever possible the methods of experimental design obtained in one field to other fields.

The "Institutes of Experimental Method" should in general consist of three sections: (1) a general methodology section whose aim will be to determine the general conditions necessary for any meaningful experiment, (2) a section on mathematical statistics, whose aim will be the development of the mathematical theory necessary for improved experimental techniques, (3) a section on applied statistics, whose aim will be to develop methods of experimental design in accordance with the criteria set down by the philosophy group and the results obtained by the mathematical group, and to apply these methods to specific fields. The section on applications would contain representatives from each important field of experimental research whose interest lies at least in part in the design of experiment (2), (4), and (6).

Such is the direction our future research should take. This essay will have succeeded in its purpose if by its argument it has convinced the reader that research in isolation must eventually fail to be pro-

ductive, and that the indefinite advance in any one field will always depend upon the advance in all fields.

REFERENCES

1. Ackoff, R. L., "Towards an Interpretation of Contemporary Philosophy," *Philosophy of Science* (1946).
2. Churchman, C. W., and Ackoff, R. L., "Varieties of Unification," *Philosophy of Science* (1946).
3. Lenzen, V., *Procedures of Empirical Science*, Univ. of Chicago Press (1938) (in *International Encyclopedia of Unified Science*).
4. *Measurement of Consumer Interest*, edited by C. W. Churchman, R. L. Ackoff, and M. Wax, Univ. of Pennsylvania Press (1947).
5. Nagel, E., "Review of Reichenbach's 'Philosophical Foundations of Quantum Mechanics,'" *Journal of Philosophy* (1945).
6. "Proposal for an Institute of Experimental Method," in *Bulletin of the Institute of Experimental Method* (U. of Pa.) (1946).
7. Reichenbach, H., *Philosophical Foundations of Quantum Mechanics*, Univ. of California Press (1945).
8. Ruddick, C. T., "On the Contingency of Natural Law," *Monist* (1932).
9. Shewhart, W. A., *Statistical Method from the Viewpoint of Quality Control*, Graduate School, Department of Agriculture (1939).

Chapter XIV On Science, Personality, and Social Conflict

The philosophy of science, whose task is to investigate the methods by which science answers its questions, is constantly faced with a problem peculiar to its field. In the study of the scientific method of problem-solving, one has continually to recognize that within his society there is displayed a multitude of ways of solving problems that are *not* scientific, and yet do in some sense satisfy the needs of the individuals involved.

For example, although the most advanced research in scientific methodology has come to the conclusion that no question in science is ever answered finally, without risk of error, on the basis of a finite set of observations, nevertheless it is well recognized that people do make decisions about which they are absolutely confident and do not expect to change their minds at a later time. One recognizes the furniture in his home and knows how to avoid it as he walks about; he knows what foods are nourishing, what things are dangerous to health; he knows the sounds, colors, smells, that form a part of his daily existence, and the information about these sensations he takes to be wholly reliable.

And he *must* be sure about these things, or else he would find it impossible to act efficiently; he cannot even entertain the notion that there are risks involved in his decisions, for if such doubts creep in, he finds it impossible to act quickly and efficiently. The refined reflection of the scientist is not an available means in so much of individual and social planning. We must decide how to act, what laws are to be supported, what men to be elected, what products

to buy, without the aid of the scientist's precise but complicated technique of observation.

Again, scientific method has demanded that the responses to scientific questions be evaluated in terms of probabilities (the long-run risks of wrong decisions), whereas in our everyday problem-solving we cannot even begin to estimate such probabilities. We are conscious, however, of degrees of confidence in our decisions, and we use these degrees in determining courses of action. We are aware that a trip across the George Washington Bridge is much safer and quicker than a trip across the Hudson in a rowboat, though we could not express our confidence in probability terms.

If the individual and his social group are forced into using nonscientific means for solving their problems, then it becomes a problem for the philosopher of science to decide whether nonscientific methodology lies within the scope of his inquiry. From the pragmatic point of view, since all this imprecise problem-solving *must* be done in a nonscientific manner, and since in general the everyday techniques work out satisfactorily, ought we not to say that there is a more general meaning of truth within the culture than the meaning assigned by the scientist?

Our answer to this question must be a decided negative, and the motivation for a negative answer has been argued frequently enough by the philosophers of science: to demand two truth-methodologies is to confuse the purpose of all experimental science, to disunify methodology so that the autonomy of science is lost in a muddle of metaphysical and epistemological principles. From the point of view of a continued progress in science, we must regard as reactionary the attempt of the epistemologist to argue from the existence of nonscientific problem-solving to an epistemology that is more general in scope than scientific methodology.

And yet if we are so positive in negating the epistemologist's claim, we must take upon ourselves the obligation of showing the relationship between everyday problem-solving and the type of problem-solving we call scientific. The neglect of this aspect of the philosophy of science has enabled the metaphysicians to establish a firm ground upon which to construct a general epistemology not subject to the rigors of scientific criticism.

The attempt is made here to outline how nonscientific problem-solving can be studied within scientific method. In fairness to the scope of this problem, this effort should only be regarded as prolegomena to future studies by methodologists in this field.

Preliminary to this study, we make certain assumptions and definitions that are consistent with the scientist's demand for a general methodology of experimental inference.

We assume that every question raised by the individual, or his social group, is raised for some purpose; i.e., we presuppose that problem-solving is an aspect of purposive behavior.

To make this assertion precise within the framework of methodology, we must explain in experimental terms what "purpose" means in this connection, and what aspects of experimental teleology are involved in problem-solving. To review briefly the results of Chapter XII, we say that a set of elements belongs to the same teleological class if they exhibit no common structural morphology, but have a common characteristic of *production*, where production is given the technical meaning assigned in Chapter XII. Thus, the class of timepieces is a teleological class, since its elements are not describable in any common structural terms, but each produces a typical kind of response among certain biological individuals ("time-telling").[1]

It will be evident that membership in a teleological class, as thus defined, does not constitute problem-solving in any of the usually accepted senses of the term. A watch, a hammer, a pencil are not "solving problems," although each belongs to a particular teleological class. Thus *function* (membership in a teleological class) is a necessary but not sufficient condition for *purpose* (problem solving). Purposive behavior implies something ascribable to the behaving individual *as an individual.*

We have said that an individual x belongs to a teleological class if he belongs to a class of elements having no common structural morphology, but having a common production-characteristic. In the same manner, we say that a teleologically defined individual x belongs to a purposive class if within structurally similar environments he exhibits behavior patterns having no common morpho-

[1] Thus, not all teleological classes are composed of "living" elements; Singer (6) has given the additional properties sufficient to define the class of Life.

logical description, but having a common production-characteristic. Thus, as I sit in this room, I may produce ink-scratches on a piece of paper in many different ways (by using a pen, a typewriter, a brush). The environment has not changed structurally, in the sense that the environments in which I might perform the various acts belong to the same morphological class. However, the acts themselves can be made so different that the analyst will find no way in which the behavior patterns can all be described by a common structural (mechanical or physical) property. An "individual" who displays diversity of behavior having a common producer-characteristic in structurally similar environments is said to have a purpose, a purpose that is typically his, and not necessarily anyone's else. The individual so defined has "freedom" in the classical sense; such freedom is the result of our defining, to be sure, and, as one might expect, the existence of freedom in a mechanically defined nature depends on the kind of question we ask. If we consider "individuals" as mechanically or physically defined, then the questions we ask and the answers we give must not involve the concept of freedom; but if individuals are teleologically defined, and we consider their purposes in the sense of teleology and purpose defined above, then we can talk about freedom, and, more important, can measure certain aspects of a free individual's behavior, without conflicting in any way with a well-defined mechanism.

Now those behavior patterns an individual may display that have a common producer-characteristic we call his *means*, and the common characteristic we call his *end*. Relative to a certain environment, we can measure the probability that the individual will display any given type of behavior; those types which he virtually never chooses we say are *nonpotential* means. Thus within a given environment, we can define in methodological terms the event of problem-solving: it is the event in which a *teleologically defined* individual selects a *potential means* from among the available means in the environment, for the accomplishment of an *end*, the italicized terms having been defined, as above, in a manner consistent with a mechanical explanation of the event.

We have said that the means used by an individual are producers of some state, or share the same producer-characteristic. Now in

general, if x is a producer of y, then there will be elements in nature having the same morphology as x, but which are nonproducers of elements of the y-morphology. Hence, the class of x-morphology can be considered with respect to the *probability* of a y-production. A morphologically similar set of light blows on a certain detonator cap will produce but few explosions, whereas a set of heavy blows will produce explosions with high frequency. Consequently, the x-morphology can be measured with respect to the probability that any random member will produce a y; we call this measure the *efficiency* of x with respect to y-production. The efficiency will evidently be a function of the environment.

We can now consider the problem-solver with respect to the *efficiency* of the *means* he uses to accomplish certain *ends*. We have thus arrived at a general technique of describing the individual with respect to particular situations in which we may want to study him. We can evaluate scientifically the methodology an individual displays, and can explain in terms of purpose the behavior patterns he exhibits.

But though we can thus reduce behavior to purposive terms, it does look as though we have not accomplished thereby a complete "explanation" of the way in which a person solves his problems. If everyone chose the most efficient means he knew about, then our scheme would be complete: the concepts of science and individual problem-solving would suffice to account for all behavior. But people do choose means that are not efficient even though they "know better." More exactly, we can examine an individual's behavior and determine the range of his *potential* means (means he will sometimes select) relative to a given environment; and we will also observe that in many cases the individual does *not* choose the most efficient means in this range.

It would not be too great a departure from traditional concepts to define "personality" as the measure (or measures) of typical inefficiency an individual displays in problem-solving. The concept of personality so defined would agree with the traditional demand that personality represent the individual *qua* individual, for it would measure his peculiar characteristics in the quest for ends that are *his*, and not necessarily those of anyone else.

Specific suggestions for an experimental definition of personality have been made elsewhere.[1] Typical measures are the measures of "lag" and "anti-lag." An individual exhibits lag when he tends to cling to old means for solving his problems, when the problem situation has so changed that other means have become more efficient; he exhibits "anti-lag" when he tends to change to new means when the old means are still the most efficient. The measures, be it noted, are measures of *tendency;* we are not interested in what an individual does in a single case, for his inefficiency might in any specific instance be due to physical or physiological influences that have nothing to do with his general pattern of behavior. We are interested in whether lag or anti-lag is typical of a wide diversity of problem situations. Hence, we are interested in measuring the probability that inefficiency of a certain type will occur.

The problem of personality cannot be said to have been completely solved until we have explained why an individual typically chooses inefficient means. Simply to measure personality is not to explain it; we must also show the origins of the measures we have obtained.

"Explanation" is a term itself that requires some explanation if we are to take the next step in the analysis of individual problem-solving. We can "explain" the behavior of the planets by stating the general laws governing their motion. Such explanation we call "mechanical," and in general these explanations are deterministic. In the same manner, it is at least theoretically possible to explain mechanically the behavior of an individual, in terms of the physical and chemical conditions of his body and its environment. Such an

[1] See (1) and (2) In (2), personality is considered as the complete description of the probability of behavior-choices of an individual, given a physical or teleological description of the environment, and the ends he may be seeking. The important measures appear to be: *familiarity,* which is the probability that an individual will choose a given means, when the alternative means for an end all have equal efficiency; *knowledge,* which is the rate of change in the probability of choice of a means, as the efficiency changes; *intention,* which is the probability that an individual will choose a means to an end in a situation where the alternative modes of behavior are each efficient for some end, but have no efficiency for the other ends (i.e., the individual has equal opportunity of pursuing any end, but can only pursue one at a time); *vacillation,* which is the rate of change in intention relative to the efficiency of the means for an end. These basic measures become the tools for defining the well-known psychological scales: degree and intensity of sensation, awareness, consciousness, belief, intelligence, conflict, character, etc.

explanation would consist of a set of physical laws under which we would subsume the particular behavior he exhibits, and, at least within classical mechanics, these laws would be deterministic. But such an explanation of individual behavior would not be complete, in the sense that it does not supply answers to all questions concerning the "individual." In so far as the individual is mechanically defined as a set of point-masses individuated in space-time, the explanation is complete; but in so far as the individual is teleologically defined as a member of a class of individuals having no common structural property, but having a common end, then the mechanical explanation is incomplete. For in the latter case the "individual" is not a set of mechanically defined points; indeed, at two different moments of time, the teleological individual may be composed of distinct sets of mass-points. The identification of a functioning individual may depend upon mechanical laws, but mechanism alone is not sufficient.

Singer has put the matter as follows:

A phrase of Heinrich Hertz comes back to remind me how tenacious on the most disillusioned intellect is the hold of primitive habits of thought. Before him a set of equations deducible from the Hamiltonian principle and certain conditions laid on a sphere rolling on a plane, Hertz cannot repress the feeling that a thing so moving would "decidedly have the appearance of a living thing, steering its course consciously towards a given goal" (3, Intr.). For a moment his mind's eye following the motions forgets its science, sees in their intricacy the inexplicable, and in the inexplicable, life. The next, his understanding has recovered its equations; they drive inexplicability before them, and with it, life. . . .

Hertz's sphere moving on a plane did, indeed, lack all spontaneity. But it lacked this not because each of its turns and twists fell into a class of events having a mechanical equation in common; it lacked it because with all its turning and twisting *it fell into no second class* having not mechanism but purpose in common. Whereas the simplest thing that lives enjoys the most objective freedom. But it enjoys this not because its turns and twists must forever baffle mechanical insight; it enjoys it because *it does fall into a second class* having purpose, not mechanism in common. Not for a moment of scientific darkness but for all the ages of scientific illumination are the phenomena of life beyond the reach of *mechanical* explanation; but they are known to be so not a moment before our science has brought them within the grasp of *teleological* explanation (5, pp. 423–424).

In sum, the same natural "event" is susceptible to more than one "explanation." The guillotine that falls and neatly severs the neck can be "explained" by the laws of falling bodies; it can also be "explained" by demonstrating its inclusion in a class of objects having no common morphology, but having a common *function*, the production of a rather unpleasant physical state; finally, it can be explained as an aspect of the behavior of an individual or group, such behavior belonging to a class of diverse behavior patterns having the common property or *purpose* of removing a dangerous element.

All these explanations of the same behavior, the *mechanical*, the *functional*, the *purposive*, are consistent, be it noted, but each answers a different question about the natural world.

Now the analyst who introduces new explanations, i.e., raises new questions, must show in what sense these are justified. Why are we pushed on beyond mechanism to function, beyond function to purpose, and finally beyond purposive behavior to ultimate values? Two replies to this question are typical of present-day writing: (1) the higher forms of explanation are *convenient*, and (2) the higher forms are *necessary*. The latter viewpoint *a fortiori* justifies the former, and has been adopted here. Briefly, the techniques of answering *any* questions in science, whether they be questions of mechanism, or teleology, or sociology, require presuppositions of *efficiency:* we must have criteria of the most efficient methods of solving problems before we can give responses to any questions. Now the extent to which we go in the hierarchy of explanations will depend upon where we take the true measure of efficiency to lie. If we take efficiency to be a term that is relative to the individual in a specific situation, then his techniques of problem-solving will be the measure of efficiency. In this narrow sense of relativism, there would be no way of distinguishing between personalities, since everyone would use the most efficient means by definition. The difficulties involved in such a narrowly conceived methodology would be that we would have no way of judging what an individual's purposes *really* were in any given situation. The analysis of "meaning" contained in Chapter XI has shown that the meaningfulness of *any* question the scientist raises depends upon a purposive analysis; but if all decisions are made relative to the indi-

vidual, no objective test would exist to determine the ends that are common to the means the scientist chooses. Each man would judge his own means and the means of others from his own point of view, but exactly what his point of view is we would not be able to say.

To escape the scepticism of this sort of analysis, it is necessary to find another ground for defining efficiency, and another type of explanation for the inefficiency we then find in individual behavior.

This higher ground we assume to exist within the social group. The concept of a social group, like the concept of personality, has been given a multitude of definitions in the literature, but the discussion of the problem of inefficiency will motivate us to ascribe certain properties to social groups. We must be able to explain individual inefficiency, for example, as an aspect of social groups. We must also incorporate into the meaning of group some kind of unifying action on the part of the members. As a provisional definition, we offer the following: a class of purposive individuals is said to form a *social group* if

1. They all have at least one common purpose Y.

We say that all means to Y are "probable-producers of Y" if the relative frequency of success is greater than one-half. When an individual's behavior can be characterized over a period of time as having the purpose of attaining such a probable-producer of Y, we say that he is pursuing a "group-goal."

2. For any member of the class, there is a positive correlation between an increase in his chances of attaining his group-goals, and the chances of any other random member's attaining *his* group-goals.

Condition 2 specifies a degree of cooperation among members of the group. It does not follow that perfect harmony must exist in the group in the pursuit of Y; such harmony would only occur when the correlation was perfect (i.e., was unity). We are merely asserting that an individual who *tends* to conflict with the aims of members of the class is not to be regarded as a member of the group. In a sense, this condition binds the group-members together; for example, a class of persons waiting to board a train might be said to have the same purpose and hence would satisfy condition 1; but they could hardly be said to form a social group, for the progress of one in

attaining his goal would not in general be related to the progress of another.

We require, finally, the aspect of inefficiency which characterizes groups.

3. There exist at least certain members of the group who have purposes relative to which Y is an inefficient means; such members, in so far as they pursue Y, exhibit "personality traits" with respect to their individual purposes that conflict with Y.

The motivation for insisting upon condition 3, which implies the existence of "sacrifice" as a criterion of the social group, could only be explained by a much more thorough review of the history of the concept than can be given here (4). Condition 3 implies that the elements of the group will show "lag" in the pursuit of their ends, and such lag, at this level of explanation, can be called "cultural lag." The condition is more general, however, than a mere specification of cultural lag, in that the members of a social group may also be expected to exhibit "anti-lag," i.e., to show tendencies to change because of the group purpose Y, when change is less efficient with respect to their individual purposes.

We are thus in a position to "explain" at least certain aspects of personality. And it is to be emphasized that such explanation is necessary in the study of cultural lag and antilag; it is poor analysis to show the existence of lag, and to "explain" this characteristic of a society as a "tendency to cling to old ways." Such explanations are tautological in that they merely repeat the meaning of the concept, and cast no light on its origin or the basis for its cure. We say that in so far as an individual is considered to be a member of a social group, his personality traits may be a consequence of his pursuit of the group-end instead of his own individual ends. For example, the group may impose an inefficient form of dress on its members, and hence those members who conform to the group style will exhibit "lag" since another type of dress would be a more efficient means for individual comfort. The group purpose in this case might be the maintainance of respect for a certain class of privileged individuals, and a difference in dress among the members of the group might help to symbolize the distinction between the exploiters and the exploited. Thus an individual who continues to

be a member of his group will have to sacrifice the pursuit of certain individual aims in the pursuit of the group goal. Or, conformity with a group purpose of imposing an "economy of scarcity" may make a man exhibit "antilag" in the form of stopping his education, or seeking another position, when either of these changes actually is inefficient with respect to his end of knowledge or wealth.

It is to be emphasized again that the sociological explanation of personality traits is not the only possible explanation. A given degree of personality might be ascribed to physiological characteristics of the individual, in which case the explanation would be reduced to biological terms. Sociological explanations take place when we can correlate the personality measure with the aims of the social group.

Returning now to condition 2 for the existence of a social group, we say that the correlation between an increase in x's chances of attaining x's group-goals and the chances of any random y's attaining *his* group-goals is a measure of *cooperation* within the social group. (Subsequent experiment might well necessitate a multiple correlation coefficient, rather than this simplified measure, but this result would not alter the significance of this discussion.) If the cooperation measure is (approximately) zero, then the individual acts *independently* of the manifold. If the correlation is negative, then the individual acts *at variance* with the group. A correlation of unity would mean a perfectly cooperative individual relative to the group.

With respect to condition 3, we say that the degree of inefficiency that the pursuit of the group purpose imposes on an individual with respect to a given end is a measure of group sacrifice. More precisely, group sacrifice is the difference between the chance of success of an individual's end if he uses the most efficient known means, and the chance of success if he pursues the group purpose. If this difference is one (maximum), then complete sacrifice exists; if the individual selects the most efficient means, he will attain a desired end, whereas if he pursues the group purpose, he will certainly fail to attain his end. For example, a "suicide squad" in war would represent such a complete group sacrifice. It is to be noted that sacrifice is relative to (1) an individual's knowledge (what means does he know about?),

(2) the group purpose, and (3) a particular end of the individual. Thus a man may sacrifice his end of self-preservation, while he furthers his end of glory, in the pursuit of a group goal.

It is further to be noted that group sacrifice is a type of "conflict," a conflict between individual and group. The term sacrifice has been used since it connotes an unsatisfactory state, but the "blame" may lie with either the individual, or the group (or possibly with neither). The concept of blame at this level demands some explanation. We have measured individual blame in terms of inefficiency in accomplishing individual goals; that is, a person acts "badly" with respect to some end if he chooses an inefficient means. Under what conditions do we assert that a social conflict is "bad," and to what aspect of the conflict do we attach the blame? This question is a part of the general question: can social conflict be explained within science?

What we now require is a measure of efficiency that will be independent of the purposes of a social group. This type of measure appears to be demanded if we are to recognize sacrifice or conflict at all within society. For if the most efficient pursuit of the group purposes is a general measure of efficient behavior, then no one is ever "sacrificed," since he will always be acting in the best possible manner when he serves the group. By definition, he who acts to serve the group goal acts most efficiently, regardless of what his other purposes might be.

We take an historical study of the group concept to imply that this type of relativism does not accord with the meaning of social group. That social groups *are* characterized by inefficiency and conflict is the lesson history forces upon us. Hence we must look for a meaning of efficiency beyond the purposes of the social group itself.

The problem we have now to consider is a very general one for all methodology. We are attempting to discover the basis for deciding that one type of action is more efficient than another; and, since the discovery of the facts and laws of science is a type of action designed for a specific purpose, we are actually attempting to characterize the "best" techniques for experimental inference from observations.

Let us look at the matter in another way. There was a time in the history of our attempts to understand science and its methods,

when scientists regarded certain of their results as "objective" facts and theories, the truths of which were independent of any ethical or value judgments. The scientist, according to this earlier phase of understanding, merely *described* the observed events of the natural world. Such a viewpoint could raise seriously Hume's problem of whether or not such a description of the past could legitimately be used as a basis for predicting the future; for this viewpoint held the past to be fixed and unalterable, the future to be variable and unpredictable. But to our present pragmatic way of thinking, *all* scientific results, whether of the past or the future, are simply *means* to ends; the ends in question may be individual goals, in which case we say that the results have psychological value; or they may be group goals, in which case the results have social value. Our general viewpoint is that the results of science have an even more general value, beyond the social group, a value we may call "ultimate" or "intrinsic," provided we do not exclude experimental measurement by the use of such terms. To such a viewpoint, assertions about the "past" and the "future" are both means to the more ultimate end we are seeking. Hume's problem arises, like most of the empiricist antinomies, out of an inadequate methodology. Take the method of science to imply a "given," the content of which is independent of purpose or value, and Hume's antinomy follows. The "past" is no more "predictable" than the future, in the sense that what we take the past to be may or may not serve the ultimate end.

This pragmatic viewpoint has been made very clear in the field of mathematical statistics. The statistician has shown that on the basis of a set of observations there are an indefinite number of ways we can draw an inference, each way entailing in general a different risk of error.[1]

But to evaluate a given method of inference, it is not enough simply to describe the risks of error. We must also know how *important* an error can be. The importance of error we will call the associated "cost" or "loss." Without entering into the technicalities of the use of the loss function, we can illustrate its meaning by a few simple examples. Compare the following three questions. What

[1] See Chapter XV.

is the (mean) lethal concentration of a dosage X (relative to a specific organism)? What is the purity of a certain compound to be used in match heads? What is the distance between the centers of gravity of two planets at time t_0? The procedures of modern experimental science tell us how to collect observations pertinent to each of these questions, provided we make clear our purposes. The first two questions are "practical" in the sense that the responses to them will be used as means to the attainable goals of the present; the last is "theoretical" in the sense that its response is a step in the progress toward an unattainable ideal. Now the techniques of inference from observation might vary for all three of these problems, in the sense that no single statistical method of analysis is "good" in all three cases. The reason for this depends upon the losses incurred by a wrong decision in each of the three cases. In selecting a hypothesis that the true lethal dosage lies in a certain interval, we are running a risk of both underestimation and overestimation. From the point of view of loss, there can be no doubt that a loss incurred from overestimation is much greater than a loss incurred from underestimation. If I say that the true lethal dosage lies in the interval 10 per cent to 15 per cent, and the *actual* value is 8 per cent, then my response, if it is adopted generally, may result in a large cost to the social group. That is, the group whose behavior is recognizable as accepting my response will act very inefficiently with respect to its ends. But if the *actual* lethal dosage is 17 per cent, then the acceptance of my response, though inefficient, will not (in general) be disastrous to the social or individual goal. Hence, the statistical method that *ought* to be employed in estimating the interval for the lethal dose from a set of observations is a method that will be heavily weighted against overestimation.

In the second example, which deals with the purity of a compound for match heads, the opposite occurs. Here an erroneous underestimation is more serious than an overestimation; if I say the impurity lies in the interval 1 per cent to 2 per cent, and it is *actually* 5 per cent, the match may fail to function; whereas, if it is actually .1 per cent, the consumer loss, at least, will not be great. But in this case the importance of overestimation and underestimation are not nearly so great as they were for overestimation of lethal doses. Hence,

the method of inference *ought* to be weighted against underestimation of impurity, but not nearly as heavily as in the first example.

Finally, what are the consequences of a wrong estimation in theoretical science? The possible losses we incur in this case will be either the abandonment of a theory that can still serve the purpose of science best, or else the continued acceptance of a theory whose purpose is outworn. The measure of these losses will depend upon the difficulties of devising alternative theories, and the complications of accepting a wrong theory from the point of view of other fields of investigation. In any case, the estimation of loss will be a much more involved task than in either of the first two examples, because the ultimate end is an ideal and not an attainable goal. The weighting of wrong answers will depend upon much more general criteria.

The result of this discussion is that the so-called objective "facts" and "theories" of the scientist depend for their confirmation upon the estimation of losses, and these losses are a function of the individual and social purposes. But the argument of the first part of this chapter has shown that the measure of loss, or, what is the same thing, the measure of efficiency, cannot depend solely upon individual or social purposes. To summarize:

1. Because of conflict of ends within an individual, efficiency cannot be a psychological measure.

2. Because of conflict of ends within the social group, efficiency cannot be a group measure.

3. Efficiency and its correlative, loss, are presuppositions of science in stating its answers to questions of fact or theory.

Hence, science demands a science of efficiency, and cannot establish such a theory within psychology or the science of social groups. The science of ethics, for such we call the measure of loss, must on the one hand belong to experimental science, and yet not be an aspect of any of the special disciplines now recognized. The science of ethics must, in other words, be subject to all the checks experimental science imposes on its fields of investigation; it must be subject to all the conditions of experimental control. At the same time, the science of ethics will have pervasive influence throughout all the sciences, from logic to sociology.

Our next endeavor must be to characterize the field of the science of ethics with respect to other fields of investigation, and then to suggest, as a beginning to the construction of such a science, how a general measurement of efficiency can be defined in experimental terms.

REFERENCES

1. Churchman, C. W., and Ackoff, R. L., "An Experimental Definition of Personality," *Philosophy of Science* (1947).
2. Churchman, C. W., and Ackoff, R. L., *Psychologistics* (mimeographed), Phila. (1947).
3. Hertz, H., *The Principles of Mechanics*, translated by D. E. Jones and J. T. Walley, London (1899).
4. Hussong, A. M., *An Analysis of the Group Concept*, Philadelphia (1931).
5. Singer, E. A., Jr., "On Spontaneity," *Journal of Philosophy* (1925).
6. Singer, E. A., Jr., *Experience and Reflection* (to appear).

Chapter XV On Chance, Loss, and Risk

The discussion of the previous chapter has taken us to the point where we must look for a measure of efficiency in a science whose scope lies beyond the science of social groups. The natural suggestion, and the one that we shall follow, is that the field of this new science must lie in the "grouping of groups"; the material of study of the science of efficiency is now taken to be the general purposes common to all the societies of the ages, if such exist. "Efficiency" then is to be measured relative to such general purposes, and not to the specific aims of individuals or social groups.

Just how the study of such universal purpose would be carried out will not be a matter of discussion here. Our aim for the present is devoted to proposing an hypothesis for an adequate measure of efficiency. The hypothesis we are to propose, to be "confirmed" in any sense, would have to be subjected for examination to an adequately designed method of history. This remark must be regarded as an important prelude to the subsequent treatment. The proposed measure of efficiency will not receive its validation through the fact that it may "appeal" to the rationally minded individual, or seem to some to make the only possible sense out of the situation. Such appeals are rationalistic in their method, and by implication attempt to set up a nonexperimental criterion for the science of ethics. It is important therefore to indicate that the principles of the science of ethics which we are about to propose must eventually be subjected to an historical test to determine their validity.

The method of inference in science, which has been discussed in the earlier chapters, demands that in responding to any question, alternative hypotheses be constructed, which provide specific

information concerning the manner in which the observations are to be collected, concerning the distribution function of the observations, and concerning the methods of selecting one of the alternatives. Now the *best* method for selecting alternatives will involve three concepts, which will be made the basis for the discussion of efficiency. These were suggested by Wald (8) and Brookner (2), and two of them were mentioned in the last chapter. The present discussion attempts to carry the problem beyond the point where Wald leaves it. The three concepts are:

1. *Chance.* Associated with any given hypothesis and method of selection, there is the long-run chance of rejecting the hypothesis when it is true; in the terms of the last chapter, this comment is translated into the assertion: associated with any problem-situation in which there are alternative means for the attainment of an end, and in which a specific technique is used for selecting one of the means, there is a long-run chance of not accepting the most efficient means.

Short of omnipotence, we cannot avoid a chance of error with respect to a technique of selecting one of the alternative hypotheses (possible means). As a presupposition to our inquiry, it does not seem contrary to our fundamental motives to insist that the evaluation of the best method of selection depends *partially* upon the associated risks. As the subsequent discussion will show, this dependence is only partial; minimizing chances of error is only *one* aspect of the best methods of selecting hypotheses.

In order to keep the three concepts which we are now considering clearly in mind, we shall introduce a simple example, the generalization of which to *any* type of question in science will be clear enough. Suppose we consider a factory manufacturing ball bearings for use in a certain machine. The *end* here is to make bearings with minimum expense which will operate "correctly" in the machine. When a lot of bearings is presented for inspection, it may be sampled, or it may be entirely examined. In either case, the method of examination is designed to select one of two alternative hypotheses: the lot is acceptable for use, or the lot is not acceptable and should be re-examined or discarded. The selection of either of these hypotheses may be regarded as a means for a certain end; the means will be the

best means if the original aim of minimum cost and satisfactory performance is attained. Now either inspection plan, sample or 100 per cent, will lead to chances of error, though the calculation of the chances will actually differ. For the sake of illustration, we will suppose that a sampling method is used; in particular, we suppose that the manufacturer segregates lots of about ten thousand, and samples one hundred from each lot. To make the inspection plan very simple, we will also suppose that he accepts the entire lot if there are no defects in the sample; otherwise he rejects. If we can presuppose random sampling in each lot inspected, then we can completely characterize the chances of error of the inspection plan with respect to defective articles. For any given lot, there are a multitude of possible assertions we can make; these all have the general form "the fraction defective in the lot is p," where p takes on values in the closed interval 0 to 1. Actually, since the lot is finite, the possible values of p are also finite, but it can be shown that it matters little with respect to our present purposes if we regard the lot to be infinite, and p to represent the *probability* of drawing a defect, since the lot is so much larger than the sample. Now, *under the presupposition of randomness*, we can determine the chance of accepting the lot for every possible value of p; the calculation in this case is very simple, for it depends upon calculating the chance that a random event (no defect) will occur one hundred times in a row when the probability that it occurs on any trial is p. The resulting function, which gives the chance of accepting the lot for every possible alternative, i.e., for every possible p-value, is called the "operating characteristic function." Such a function completely characterizes the chances of error associated with any given inspection plan, for it tells us how often we will reject (or accept) a given hypothesis for every possible form nature could take on.[1]

It is to be noted in passing that though the example given is simple, and the operating characteristic function contains only one variable (p), the entire procedure can be generalized to very complicated experimental tests. Suppose we wish to determine whether a certain particle travels over a path specified by theory, and design

[1] See (1), (5), (6), and (9) In addition, there are a great many other sampling plans to be found in the literature.

a certain technique for "analyzing" the data. The path may be described by n parameters, each estimated from the data with different standard errors and/or other measures of dispersion. Even in this case, relative to a given technique of analysis, it is possible to set up an operating characteristic function, with respect to the chance of accepting the hypothesis that the particle does in fact travel the theoretical path. Such a function would give the chance of accepting the hypothesis for all possible values of the parameters and their errors of estimate; as before, this operating characteristic function can only be calculated provided certain presuppositions are made about the observations.

2. *Loss.* It will be evident that the chance of accepting an hypothesis when it is false is not a sufficient basis for judging the most efficient method of selecting hypotheses. If the chance of error alone were the sole basis for evaluating methods of inference, then we would never reach any decision, but would merely keep increasing the sample size indefinitely. Besides the chance of error, we must evaluate the "loss" or "cost" of a decision. Loss, or cost, in this connection has a very general meaning, not necessarily measurable by monetary standards; loss could arise in societies which had no monetary system. The manner of representing loss follows much the same pattern as the method of representing chance: we can set a "loss-operating function," which describes the loss in accepting the hypothesis for any given "state" of nature. Returning to the original example, we can give a function which states the loss in accepting the lot of bearings for any value of p, the probability of a defective. Though we shall want later to discuss in some detail the methods of measuring loss, we can for the sake of this illustration proceed in an intuitive manner. We can say that for values of p from 0 up to a certain amount, say .001, the loss in accepting the lot on the first sample is negligible, in the sense that practically uniform efficiency of operation will result from all such lots. Beyond this value, the loss will begin to rise, and finally for a certain value of p it may become very great, e.g., when the frequency of defective bearings results in disastrous breakdowns of machinery. In some cases the loss may actually become indefinitely large; for example, if the test is to decide whether a widely used food is free of a certain

deadly bacteria, then the loss from a wrong decision may amount to many lives. Or the test might be to decide whether an atomic explosion will develop a chain reaction, and, more to the point, whether the development of atomic energy will lead to widespread destruction through social misuse. Science is obligated to make estimates of all potential losses.

Now in general, the test procedure should be designed so that as the loss curve rises, the chance of error decreases. That is, as the loss resulting from a wrong decision increases, the chance of accepting the decision decreases. This demand is generally referred to as a demand for a lack of bias, a biased test being one in which, at least for certain values of the quantities under test, the chance of error does not decrease as the loss increases. This is a generalization of the definition of bias given on p. 28.

3. *Risk.* The concept of bias suggests that the value of any test procedure depends upon a certain function of both the chance of error and the loss. Associated with every possible natural state, relative to a certain problem and its presuppositions, there will be a long-run chance of accepting an hypothesis, and the loss involved in acceptance; the risk will be some function of these two quantities. Thus, in the original example, the risk curve might have the following properties: for low values of p, where the loss of acceptance is negligible, even though the chance of acceptance is high, the risk will also be negligible. For that value of p at which the loss begins to rise, the risk will also rise, but since the chance of error is decreasing, the risk will not rise as sharply as the loss. And finally, for those values at which the loss is excessively high, the risk may again be negligible, since the chance of error will be extremely small.

On the basis of the meanings of the concepts chance, loss, and risk, the following rather obvious demands of any adequate method of selecting hypotheses can be made:

1. The total risk should be finite. This means that in cases where the loss rises critically for a wrong decision (e.g., in questions concerning items affecting life or health) the chance of a wrong decision should rapidly approach zero. This can be effectuated by an increased sample size, and by introducing various techniques of analysis designed to make the most efficient use of presuppositions.

2. All other aspects being equal, that method of inference is "best" for which the total risk is at a minimum. If the loss can be represented as a continuous function of possible values of the unknown parameters, the total risk will be an integral if such exists; where the risk is discontinuous, or where the parameters take on discrete values, the total risk will be a finite summation.

3. All other aspects being equal, the method of inference is "best" for which the maximum of the risk curve is a minimum.

These are but a few of the desirable qualities of a risk curve. The subsequent study will be designed to determine the general criteria of value that not only enable one to determine the best method of inference, but to measure the loss and to set up the risk function.

It is interesting to note in passing how the basic concepts of chance, loss, and risk are fundamental to the concept of meaning. As a critical case, let us suppose that a question is asked such that any answer involves no loss; then it seems obvious enough to say that the risk involved in any response to the question is always zero. But such a question is also meaningless, for we shall certainly want to say that the loss depends upon the efficiency of a response with respect to certain purposes, and if all behavior patterns are equally "good" with respect to a goal, then no purpose exists. But by the criteria already set down in the earlier chapters, questions which are raised without a purpose are meaningless.

On the other hand, let us suppose that an ideal situation has been reached in which the chance of a wrong decision has been reduced to zero. Then again it seems evident that the risk will also be zero. Thus both meaningless questions, and questions which can be ideally answered (i.e., without chance of error) are questions involving no loss; the meaningless questions have associated with them a zero loss, the answered questions have associated with them a zero chance of error.

The fundamental task is to determine the criteria of loss. Evidenly, these criteria are not fundamentally statistical in nature, so that the criteria of correct inference from data do not lie solely within statistical research.

In order to keep the problem in a clear focus, let us suppose that we are interested in measuring the loss in submitting a certain con-

sumer good to public consumption. The empirical situation can be described as follows: the article is made available to an individual, who intends to use it for a certain purpose. We are now interested in measuring the loss to such a consumer if he follows out his intentions. For example, a certain kind of tooth paste is put on the market; in evaluating the product with respect to the consumer's interest, we would like to know the possible losses he would incur in using this tooth paste rather than any other.

Now one difficulty with this simplified picture of things will be immediately apparent. An individual, or group, has a wide number of purposes, some of which conflict with others. Hence, the use of an article may incur a large loss with respect to one purpose, and no loss at all with respect to another. For example, drinking poor coffee, while it satisfies the purpose of stimulation, may be detrimental to the purpose of good health.

The preliminary task will be to set up criteria for measuring loss with respect to a specific purpose, and then attempt to generalize on loss measurements over all the manifold of purposes exhibited by a social group.

Associated with any given purpose of an individual or group, there will be the probability of the purpose's being fulfilled if the article is used in a certain manner under certain environmental conditions. The problem of measuring such probability is simply the problem of evaluating an article with respect to the concept of "probable-producer" in a specific environment. We now define the "legitimacy" of the consumer good:

The measure of the legitimacy of a consumer good X *with respect to a specific purpose* Y *of an individual or group, within a specific environment, is the ratio of the probability of attaining* Y *if* X *is used, to the probability of attaining* Y *if the most efficient known means is used.*[1] It is to be emphasized that the definition of legitimacy is restricted to (a) specific goals, (b) specific users, (c) specific environmental conditions, and (d) the state of knowledge of the group. From the point of view of the aims we have in mind, there is also involved a circularity, for if legitimacy is to be used as a basis for measuring loss, and loss to be used as a basis for evaluating inference, then the

[1] Thus the "legitimacy" is the relative efficiency; it ranges from 0 to 1.

adequacy of measures which are presupposed in measuring legitimacy will depend upon the concept itself.

We shall want to return to this inevitable aspect of circularity later. For the present, we should make some comment about the term "knowledge" which is involved in the measurement of legitimacy. We suppose that it is experimentally possible for the psychologist, or social psychologist, to determine for any given purpose and environment what means an individual or group *could* choose with probability greater than a virtual zero. We say that the range of means that could be chosen, as measured by such tests, constitutes the *knowledge* of the individual or group. The measure of the legitimacy of an article is, therefore, relative to knowledge; we are not requiring the impossible task of comparing the article with the most efficient possible means, for, short of omnipotence, such means are not determinable. But we can objectively determine what means are known to a group, and compare, with respect to the probable production of the desired end, the efficiency of any article with the most efficient known means.

To illustrate, a manufacturer places on the market a soap which sells for the same price as certain other brands. The "goal" Y may be specified by certain laboratory tests: the soap is to cleanse a type of cloth, containing specified amounts of organic and inorganic impurities, to a certain degree of purity within a particular time-interval. Random samples of the soap are selected, and the test is conducted to determine the probability of success. The difference between this probability, and the probability for the best soap that could be made, will be the measure of the legitimacy of the new soap.

Now the goal Y in the example above is so specific in meaning that the legitimacy score would actually have very little practical value. In most cases it will be possible to specify more general purposes, so that the laboratory tests are conducted over a range of the variables involved, and so that the probability becomes a function of certain variables such as time, temperature of the water, and so on. But the purpose should not be so generalized that it becomes impossible to compare two articles; for example, one soap might cleanse more quickly but to a lesser degree of purity than another.

If this were the case, then the soaps would have to be scored with respect to legitimacy against two different purposes, rather than the one purpose of cleanliness in the shortest possible time.

It is now necessary to generalize upon the measure of legitimacy, so that we can evaluate means with respect to all purposes. The various legitimacy measures must be "weighted" in accordance with the importance of the associated purpose. In view of the previous discussion of personality and social groups, the determination of these weights appears impossible. The same individual has conflicting purposes; the same social group desires goals which cannot be simultaneously fulfilled. Consequently, we cannot weight the legitimacy scores on the hedonistic basis of satisfaction of individual desires, or the liberal basis of the satisfaction of social desires. The fallacy of both egoistic hedonism and relativistic liberalism lies in ignoring the fact that there exist a set of conflicting aims within each individual or social group. The "happiness" or "welfare" of an individual does not depend only on examination of his own aims; these aims in general are in conflict, so that the "good" cannot be individualized. Neither can the happiness or welfare of the social group be determined from an examination of the aims of the group; all groups are exploitive in the sense that membership in the group involves some cost with respect to the aims of at least some of its members. The liberal's plea for governmental policies based upon the common welfare are, therefore, misconceived. Thus, in a time of severe social strife between labor and capital, the liberal is apt to frown on what he conceives to be indiscriminate striking, or on strikes which cripple the national economy, since he asserts that such activities are against the "common welfare." And yet, within such social groups, there can be no common welfare, since the existence of the conflict, and its generality, preclude the possibility of *one* aim which has high importance for all, or even a large part of the membership.

Now the basis of the fallacy of the liberalism which seeks for a unified aim within a specific society is that it demands an "explanation" of group goals without generalizing upon the nature of groups. We have already seen how the teleologist generalizes upon the purely mechanical explanation of an act, so that the act can also be

interpreted as purposeful; we have also seen how typically inefficient teleological behavior can receive an explanation within the purposes of the social group. The basic aspects of both personality and social group are *conflict;* the only possibility of "removing" conflict, i.e., determining purposes which transcend conflict, is to move beyond the social group. What we now require is a way of examining man's activity which will "interpret" the conflict of aims within the social group as itself serving a unified purpose. Just as the social group served to explain the conflict of aims within an individual, so the next step must explain the conflict between the elements of society.

As a first suggestion, one might attempt to define the purpose which transcends all group conflict as the purpose of satisfying the maximum number of nonconflicting desires.[1] Put in experimental terms, we might first propose that the measurement of "progress" is the measure of the random individual's "power," i.e., the measure of the probability that he will attain the maximum number of nonconflicting goals. Such a purpose would be "general": it would in some sense transcend the purpose of individuals and social groups.

But such a formulation does not avoid one aspect of progress that seems to be demanded from a common-sense level. We do want the ideal which defines progress to be free of any "exploitive" tinge. For one thing, we wish to have no sympathy with a program aimed to simplify our desires, or eliminate them. One of the most insidious types of social exploitation occurs when the preferences of a group are perverted and simplified, so that actually the members of the exploited group desire very little, or desire only objects their exploiters can cheaply and profitably supply. We could hardly assign a very high legitimacy score to a purpose of an exploiter who was trying to pervert the group desires in this manner, even though he is actually working to satisfy the maximum number of *existent* goals of a group. In other words, "power" alone is not a sufficient criterion for the ideal or generalized purpose in terms of which all specialized aims are to be evaluated. For power may actually increase as desires decrease, so that in one sense the desire-less individual is most powerful. The most extreme exploiter, under this conception, would be the most beneficent member of mankind.

[1] See (7).

Admitting as inadequate this simple definition of a most general purpose (one satisfying the maximum number of nonconflicting desires), we must proceed to a refinement. The discussion so far has shown that there are (at least) two fundamental requirements of the ideal which is to serve as the basis for measuring progress and for measuring the general legitimacy of specialized goals:

1. The ideal should in some sense be independent of the conflict of purposes which characterizes the individual and the social group.
2. The ideal should not be consistent with a minimization of desires.

To make a long story short and oversimplified, we suppose that it is possible by an examination of the histories of societies with respect to their aims and conflicts, to determine *predominant* purposes expressive of the aims of man, not as viewed from one age or social group, but as viewed throughout all the changes of societies in their various historical aspects. Such predominant purposes let us call "historical." Let us then define the most general purpose, or ideal, to be the satisfaction of any given historical purpose; or, in experimental terms, let us say that the measure of progress is the measure of a random individual's power (probability of attainment) with respect to the set of historical purposes. Examples of such historical purposes would evidently be health, comfort, security, and similar aims.

The concept of historical purposes or aims may seem to be tainted with the nonexperimental brush we have so often discarded in this essay. To make the point of view clear, we are demanding an experimental science of history as a basis for determining the predominant purposes. The program we are proposing is certainly a large one; it entails setting up a science of history adequate to the discovery of predominant purposes inherent in the conflict of aims peculiar to any given society or culture, and a method of evaluating specific ends in terms of their efficiency with respect to the general end. The formal background of such an historical science would presuppose that predominant purposes are determinable, *not* in spite of conflicts within social groups, but as generalized aspects of such conflicts. That is, the conflict inherent in social groups would form a *basis* for determining predominant purposes. In a sense, the predominant

purposes would be *analogous* to the general laws of a mechanical image, which are sufficient to "explain" the particular occurrences in various "time-slices" of the image.

Although crude beginnings of legitimacy scores could even now be effectuated within social science, it will be clear that the future success of the proposed program depends upon our attitude regarding the future of methods within the social sciences. If we argue that we can never attain experimental controls in this domain, then we shall have to admit in the same breath that we can never hope to evaluate social aims. For one who wishes to generalize on these matters, to give up the hope of experimental control and historical research is tantamount to giving up the hope of any science, any experiment, any control at all; for all science must be evaluated with respect to its aims and methods. This is the reason that ethics becomes an essential science; like any other aspect of science, if we give up the possibility of making progress in this domain, we give up the possibility of making progress in any domain.

If weighted legitimacy scores can be obtained, then it should be possible to formulate measures of loss and risk, and to construct a science adequate to answer the basic problem of decision which statistical theory leaves unanswered. This is the appeal of this essay: that the sciences, and society in general, collaborate in setting up a controlled science of losses and risks, i.e., a controlled science of ethics.

It will be noted that the end of this discussion has taken us around to the beginning again; for we started the research into method by studying the manner in which certain special sciences attempt to answer their problems. This forced us into a study of individual, social, and finally historical purposes. But the study of historical purposes will depend upon the activities of men in the pursuance of their aims, and in particular, upon the activities of the special sciences. Put otherwise, the meaningfulness of any particular problem of the sciences depends upon the meaning of the ideal of all scientific activity; but the ideal itself only gains meaning through the activities of the sciences themselves. We are "better off" today in stating what are our goals in science than we were a century ago, and we are better off only because of the specialized activities of the

scientists themselves. Science must provide the data by which the philosopher evaluates the criteria of adequacy of science's methods.

REFERENCES

1. Bartky, W., "Multiple Sampling with Constant Probability," *Annals of Mathmatical Statistics* (1945).
2. Brookner, R. J., "Choice of One among Several Statistical Hypotheses," *Annals of Mathematical Statistics* (1945).
3. Churchman, C. W., "The Consumer and His Interests," in *Measurement of Consumer Interest*, Ed. by C. W. Churchman, R. L. Ackoff, and M. Wax, Univ. of Penna. Press (1947).
4. Churchman, C. W., and Ackoff, R. L., "Footnote to 'Logic of Statistical Tests,'" *Bulletin of the Institute of Experimental Method*, Vol. 1, No. 4 (1947).
5. Dodge, H. F., and Romig, H. G., "A Method of Sampling Inspection," *Bell System Technical Journal* (1929).
6. Dodge, H. F., and Romig, H. G., "Single Sampling and Double Sampling Inspection Tables," *Bell System Technical Journal* (1941).
7. Singer, E. A., Jr., *On the Contented Life*, Holt (1936).
8. Wald, A., *On the Principles of Statistical Inference*, Notre Dame (1942).
9. Wald, A., and Wolfowitz, J. W., "Sampling Inspection Plans for Continuous Production," *Annals of Mathematical Statistics* (1945).

Chapter XVI On the Control of Quality — An Ideal

In this final chapter, we shall discuss an application of the philosophy of scientific method which has been developed within this book. This application of course will entail a repetition of many of the thoughts already expressed, but the general purpose here is to show how the ideas expressed herein may enlarge our understanding of an important phase of methodological research.

An application of the philosophy of scientific method that is frequently ignored, but the importance of which is difficult to overestimate, is the problem of the control of the quality of manufactured products. We are all too prone to think of the task of science to be concerned with the answering of questions in terms of repetitive observations; the fact is often overlooked that the problem of repetitive and controlled manufacture is analogous in all respects to the problem of repetitive and controlled observation. The scientist who sets out to solve a certain problem has the following demands to satisfy:

a) The question asked must be so stated that a scientific methodology can be established for making responses to it within a prescribed degree of precision.

b) A set of observations of size n are to be taken.

c) In order to give a "satisfactory" answer to the question asked, within the required degree of precision, the experimenter must "control" the observations.

A similar set of problems faces the manufacturer:

a) Specifications must be so stated that a manufacturing process can be established for producing material within a prescribed tolerance.

b) A production procedure is to be established to produce and distribute material in a certain amount.

c) In order that the material produced and distributed be considered "satisfactory," the manufacturer must exercise control over his operations.

It is not surprising that there is such close similarity between the efforts of the experimenter and the efforts of the manufacturer. Evidently there must be some common purpose both share, a purpose that in effect is a generalization of the purpose we have previously ascribed to the scientist. The study of the methodology of controlled manufacturing should thus enable us to generalize upon the description we have already given of the methodology of controlled experiment, and to answer some questions the previous discussion left unanswered. In particular, we should be able to determine the relationship between the purpose of science (reduction of error) and the purpose of productive manufacture; for the procedures of manufacturing appear to be the direct application of the work of the scientist, and in some sense we should expect to find that progress in manufacturing techniques depends upon progress in science.

Since the treatment of the earlier chapters was couched in terms different from those listed above, it will be well to review scientific methodology under the threefold aspect given here, before proceeding to its analogue in manufacturing procedures. We will find, as suggested, that this revised treatment will take us somewhat beyond the discussions of the earlier chapters.

We have already suggested a partial answer to the problem raised under a): What are the criteria of meaning of scientific methodology? We have argued that a rationalist formalism or an empirical operationalism are not sufficient to establish general criteria. Instead, an experimentalist viewpoint has been outlined. We have demanded that pertinent observations be made, that a formal theory be constructed within which the question can be answered, that the theory be "tested" by means of the observations and a set of rules of inference supplied by statistical theory. We have suggested that the weakness of rationalism lies in its assumptions that observations can never lead to discarding a theory well established in "reason." The weakness of empiricism we took to lie in its

assumption that the pertinence of observations can be established independent of theory, by means of "definite" physical operations. The experimentalist viewpoint is that both problems, the adequacy of theory and the definiteness of observation, must also be put to the test of scientific method. In other words, a scientific method should be general enough to include the answering of these questions as well as others.

So much was argued on the side of methodology; but we also were led to discard methodology as a complete answer to the problem of science. We observed that a relativism or pragmatism that bases meaning solely on method is led to a profound scepticism that casts out forever any absolute meaningfulness to questions asked. For the relativist, the "specifications" of science enjoy a freedom from all absolute restrictions and because of this freedom bear the mark of anarchism.

Over and above methodology, we added the requirement that science have a generalized purpose, or endpoint, in the sense that no question can be said to have meaning, unless its answering could be considered to serve the scientific purpose. Progress in the accomplishment of this purpose we took to be measured by the reduction of the error of measurement. Such a formulation of criteria of meaning led us by necessity to the second and third points listed under b) and c). We found that the ideal of an errorless measurement could only be approached by taking observations in indefinitely increasing number, and that there was a constant demand for the experimenter to decide whether the ideal was being approached satisfactorily or not, i.e., whether the observations are or are not "in control." The discussion of points under b) and c) led us to regard the three points here listed as so dependent on each other for their meaning that they ought better to be regarded as the triple aspect of one unifying scientific purpose.

The problem that still remains, and carries us on to the next step, is one only briefly suggested before: granted that we now have a basis for an adequate characterization of scientific method, *what* questions are we to select for study? The treatment to date has assumed that a question of fact or law has been presented to the scientist, and has considered only the problem of how a response is

to be given. But surely it is important to decide *what* questions one is to answer, not just *how* questions are to be answered.

The problem is now concerned with the *importance* of questions. In general, one who devoted his life to counting the grains of sand on a beach would not be considered much of a scientist, even though he did keep improving his method; while a Harvey, or a Copernicus, for all their lack of method, were great scientists because in part the problems they considered were significant. If the description of scientific method is to be complete, we must decide upon criteria of importance with respect to questions asked.

To a rationalist, like Spinoza, the importance of a problem lay in its generality, since the answering of a general problem would enable one to deduce a greater amount of information than would the answering of a less general one. This criterion, which undoubtedly requires a more exact defining, is frequently employed in the pure mathematical sciences; the importance of a mathematical discovery is based, among other things, on the generality of the theorem deduced. To an empirical temperament, the importance of scientific questions is based on the collection of significant facts, facts that appear "critical," facts which lead us to abandon certain courses of action and adopt others. It requires no very keen analysis to see that neither the generalizer's (rationalist's) criterion nor the specializer's (empiricist's) are sufficient in themselves; the importance of generality does depend upon deductions we make therefrom, and these deductions usually entail more specialized information; hence the importance of generality depends upon the kinds of specialization we can derive. On the other hand, the critical facts of the empiricist are critical only because they change our attitude towards certain important courses of action or beliefs; but "courses of action" and "beliefs" are more general than the facts, and hence the importance of particular facts depends upon the kind of generalities that are involved.

To the pragmatist, the whole dispute concerning the importance of scientific questions is answered by a reference to the end or purpose of the one who asks the question. A question of importance for me may have no importance at all to you; what care you whether the pen that writes these words is blunt or sharp? In general, a

question of importance to our present culture may not be relevant to another culture at all, may indeed be meaningless. What care we what are the present mutterings of the oracle at Delphi?

We who are led to the next step beyond relativism, to the search for a unifying purpose of all science, thus find the solution to the problem of importance in this very purpose: *One question (of fact or law) is said to be more important than another if a response to it leads to a greater degree of progress*, where the degrees of progress may be measured by the general reduction of the errors of estimates of empirical quantities.

There can be no doubt that in so defining importance we have left unspecified much that is vague in our terminology. For example, it would be necessary to pick a scale for the degrees of progress which would be sensitive enough to measure progress in all fields of science. It is doubtful whether we could succeed in this task at the present point, but one sees no limitations on our indefinite approximation to this objective.

Assuming, then, that the criteria of importance, like the criteria of meaning, can be made more and more precise, we are still pushed on to the next step in the analysis: why is the purpose of science an important one for the human race? Granted that *within* science we can make a beginning to grading importance, how can we grade the importance of the whole of science in comparison with other activities?

This question, which in the present treatment of science appears last, is actually the first question which one puts in the analysis of manufacturing procedures. The "What for?" is far more apparent here, where the tradition of millennia is replaced by the experience of a century or so. But in the urgency of the demand we may find answers to both questions: the ultimate purpose of science and of production.

To one familiar with the writing of specifications for a manufactured product, it will be no new story to relate the difficulties to be faced in deciding what quality should be specified. The magnitude of these difficulties for ordinary everyday, peace-time products can be gauged by their magnitude with respect to war materiel, where the nature of the consumer's wants should be obvious. At least we

who were suddenly put in the position of writing war-time ordinance specifications thought the task would be an obvious one; we thought, for example, that all one needed to do was to decide what the manufacturer could produce, and reconcile this with what "the services wanted," and on the basis of this information write a specification which defined acceptable quality.

The trouble with this simple analysis was that the services, when one could waylay a representative long enough to ask him, said that they wanted "ammunition that would function, no more, and no less." It did no good to point out that our difficulty really lay in how to translate this request into a language within which one could define the degree of satisfaction of the consumer. But the blame for our troubles did not lie with the services themselves; one could not expect a hard-pressed gun crew to make decisions about quality "in addition to their other duties." The services rightly felt that the measure of degree of satisfaction should be studied as carefully as the measures of ballistic performance. It was as futile to trust the offhand judgments of the people who had seen action as it would be to trust someone's guess as to the hardness of a metal. In other words, it was obvious to the specification writer that in many instances he knew only half of the story; he did have good information on the kind of quality that could be produced, but he had very inexact data on the kind of quality the consumer wanted. Precision seemed to end at the door of the laboratory; information concerning the outside world and its wants was always meager.

Whether or not it would have been an easy matter to collect reliable information on the needs of the army services when such was lacking, it seems all too obvious that no such ease exists in collecting reliable information on the needs of the civilian consumer. There are essentially two problems here; one is to obtain adequate measures of what quality the consumer wants; the second is to decide what procedure is to be followed in case these wants conflict in some manner.

It has only been in the past fifty years or so that we have come to realize that the question what an individual wants can be considered as objectively as the question what a chemical compound contains.

The realization has come out of a conflict between two schools of thought on psychological methodology: the introspectionists and the behaviorists. The introspectionist thesis was that some questions regarding the desires of an individual can be answered by that individual and only by that individual. The introspectionist procedure of determining consumer interests was therefore nonmethodological, in that the response to questions of desire were thought to be an individualistic reaction and could not be generalized. That is, since all methodology supplies a general technique for answering questions, the introspectionist viewpoint implied a breakdown of method. The behaviorist went to the opposite extreme and denied the meaningfulness of questions whose answering appeared to demand the individual alone; thus, the extreme behaviorist was prone to cast out the concepts of "consciousness," "subjective desire," and such, from the realm of scientific meaning. In such action, he quite naturally called down upon himself the criticism of disregarding the historical importance of a problem in an attempt to complete a methodological system. The behaviorists would have fared better had they understood clearly the correct motive for opposing the introspectionist psychology; the motive lay not in ignoring the problems and concepts of introspectionism, but in showing how these problems could be answered by a general scientific method. The important thing is not to show that "there is no such thing as consciousness," but to show that questions regarding consciousness can be answered by experimental methods.

Now the thesis we have been developing in these pages imposes two requirements upon an adequate scientific method. We have in effect agreed with the operationalist that in order for a question to have meaning, one must be able to set up as specific a series of actions as possible, to obtain thereby a set of "pertinent" observations. We have indicated that the criteria of pertinence are "formal"; that is, the justification for assuming that a certain set of actions produces pertinent observations depends upon theoretical (formal) considerations on the part of the experimenter. These considerations must be presupposed by him in conducting his experiments. The more aware he is of the nature of these presuppositions, the more exact is his experimental method.

But the precision of operation is not enough; we have also demanded that with an increase in observation one "improve" with respect to the responses to the question asked. In order that an adequate method exist, one must be able to show that more exact answers can be given, and that the ideal of a perfect answer can be formulated as the limit of an infinite sequence of responses made according to the prescribed method.

The following demands, then, must be satisfied by any population-sampling designed to discover desires, or interests, of consumer groups: (1) A method of action should be prescribed enabling one to collect observations that are pertinent to the question asked, and to ignore results that are not pertinent; (2) It should be demonstrable that under the method prescribed, the degree of precision of responses can be increased without limit.

Thus viewed, population-sampling becomes a very complicated problem. We are evidently not going to find out how people feel about a certain product, or about a certain governmental policy, simply by asking them, unless we proceed under the very naive assumption that the answer "yes" indicates a "true" desire. Rather, we must become aware that the problem of measuring desire shares the common property of all measurement-problems: we cannot determine within a finite number of observations what a person wants at a given time, but we can make more and more precise estimates. How? The last two chapters have attempted to outline the direction such an answer should take. The criteria of adequate method require answers to questions of the following general nature: What is the relation between the verbal response to a question and the real desire? What other basis than the purely oral is available for making measurements of desires? The answering of these questions in the end will depend upon the answering of still more general questions, for evidently the relation between verbal response and true desire depends on the personality and character of the individual, and depends as well on the "personality" and "character" of the social group. Similarly, the determination of pertinent nonoral observations will depend on an adequate personality psychology and social psychology. As the previous discussion has shown, we must give up mere data gathering and statistical

analyses, and begin to formulate a general theory of personality; the theory must be precise enough to be put to the test of observation, and rich enough to enable us to make predictions concerning individual and group desires. It is as important to increase the precision and scope of such a theory as it is to collect "data"; indeed, the collection of data that have any value depends upon progress in the formal background. Though we are only beginning to develop methods of population-sampling, and to understand the difficulties, these philosophical criteria should be enough to cast doubt on some of the so-called polls of "public" opinion that are popular today.

Even though one may accept the thesis we have been developing here, that public opinion, and specifically public interests, can be measured, there still remains the question of the value of such researches. As the previous chapter has argued, it appears futile to attempt to write specifications that will be based on consumer desires, when these desires are really limitless and conflicting. Does not everyone want more automobiles, if he can have enough garages to hold them, enough gasoline and parts to keep them running? And there is no limit to this "more," is there? Worse still, Jones wants just *this* item which Smith wants as well. How can there be a general specification which will satisfy desires related to individual objects? What desires are we to satisfy in the specification of quality? Can any desire ever be satisfied?

It is difficult to understand how we can have escaped for so long a time facing this problem of value in social theory. Much of the blame is undoubtedly due to the pragmatic or instrumentalist theory, which attempts to make all value concepts relative to some purpose or aim of the culture. By facile references to "freedom" and "welfare," the exact meaning of "satisfaction" of desire, or accomplishment of purpose, is kept in abeyance.

The last two chapters have been devoted to the construction of a value-theory. We have attempted to develop general criteria of "legitimacy" with respect to consumer goods. These criteria are based on general purposes which can be determined from the history of societies, i.e., from a history of social conflict.

The criteria of legitimacy place limitations on the consumer desires that are to be satisfied at any time. This suggests a general

social program for the reduction of such limitations. If a conflict of interests exists, then sacrifice must be demanded of the members of the group, in order that the measure of cooperation be at a maximum. Such is the purpose of the state: to exact sacrifice when conflict exists. But if progress in the direction of cooperation is to be consistent with progress in the direction of satisfying all legitimate individual demands, then it must be possible to approach the ultimate (unattainable) goal of no sacrifice, and hence the abolition of the state.

The discussion so far has had the purpose of showing the possibility of writing specifications that are based on consumer interests; the thesis has been that (a) consumer desires are measurable, and (b) when conflict of desires exists, an exact basis for writing a specification can still be formulated by means of the criterion of legitimacy. What now remain are the problems of production and control.

In the spirit of this treatment, which is general in its scope, we only outline the characteristics of the solutions to these problems. How is production to be accomplished, and how can we decide that a given production process meets the demands of a specification?

Production, as we have used the term here, implies the complete manufacturing and distribution process, all the way from the shaping of raw materials to their final consumption. Once one has decided what material to make and how to distribute it (the problem of specification), one has still the social problem of action. The fundamental question here is how one gets a social group to take action: or how does one prevent curtailment of production, i.e., economy of scarcity in the presence of plenty? Such curtailment is common enough in our age, and it is futile to ascribe it to the machinations of individuals or groups of individuals. It is a social result, a collective of individual conflicts and interests. The problem here is one of writing a general specification for society, which will allow the maximum production to satisfy the interests of its members. The measure of progress in this connection is the difference between the degree to which a society can satisfy legitimate wants, and the degree to which it actually does satisfy wants. It is the measure of the *inefficiency* of a social group and in this connection is closely related

to the "personality" of an individual, since by personality we mean the typical choice of means of an individual in the accomplishment of an end. This typical choice of means is rarely based on the most efficient choice, and hence personality becomes in general a measure of inefficiency. If the analogy can be carried one step further, we might call the characteristics of the state an aspect of the "personality" of the community. The ideal here is a production and distribution process which will have maximum efficiency, and hence the ideal is really the perfection of the community personality.

In order that the efficiency of the social group be measurable, we shall have to devise methods of measuring the maximum production of the group, i.e., the maximum degree to which wants can be satisfied within the group. The difficulties in making such measures are clear enough; the measurements must be made by the state itself, and the personal bias of the group is almost certain to be reflected in the individuals who make the measurements. This means that the social group itself will hamper, and even prevent, an unbiased measure of its own efficiency. The analogue would be a psychological experiment conducted by an individual upon himself, where the only controls were such as the individual himself supplied. It might be thought that for groups within the state, the difficulty can be overcome by assigning the responsibility of measuring efficiency to "disinterested" individuals, such as universities or government bureaus. But the increase in specialization of function that is characteristic of our present civilization has made *all* groups highly dependent on one another: the activities of universities and government bureaus are not independent of the activities of industry, say. We cannot, in the face of repeated examples, deny that if a group-interest is strong enough, it can block the correct measure of its own efficiency.

Yet, though the problem of obtaining unbiased estimates of our own production potentialities is an extremely difficult one, it is also highly important, probably the most important of all we have discussed. For what value will it be to measure public interest, and remove conflict among interests, if our production is such that the degree to which we satisfy these interests is small? After all, cooperation could exist in an economy of extreme scarcity, could it not?

The moral obligation of a community is not only to remove exploitation, but also to increase without limit the probability that any random individual will satisfy all his legitimate desires, i.e., all desires that are consistent with the general aims of mankind. The situation is now analogous to the activities of science; in science, not only do we demand a method for answering questions, but we also demand that we become better and better in the answering. So here we demand of a progressive society that it not only provide a technique for evaluating and making consistent consumer desires, but that it also become better and better in its role of satisfying these desires, and that it be able to measure the degree of progress in this direction, just as science can measure the degree of her progress towards absolute precision.

We can at present only suggest how the bias in measurements of social progress is to be removed. The scientist, when faced with what appears to be an insolvable antinomy, attempts to generalize his problem on a higher level; his more general viewpoint enables him to see that what had previously appeared to be contradictories are now to be regarded as contraries, leaving room for a third possibility. The antinomy here has the following formal character: if we grant that a social group should satisfy wants in the most efficient manner, then we have on the one hand (a) the group must measure the efficiency of its production and distribution process, and on the other hand (b) the group cannot measure its own efficiency without bias.

The solution to the antinomy is found by generalizing the problem of progress. It is not necessarily the social group that progresses; it is not the wants of a collective that we seek to satisfy. It is the totalitarian fallacy to assert that by making the state into an integrated individual, and by satisfying the "desires" of this individual, we achieve the ultimate desideratum of all human wills. No, what is wanted is the satisfaction of desires of individuals taken "distributively," not collectively. And hence it is not in the social group, taken as a collection, that we are to look for progress. The moral of the antinomy is now clear: to insist on the continued maintenance of the state at all stages of history is to give up the hope of approaching the human ideal within any given distance.

The central problem now becomes one of determining the origins of totalitarianism. The demand of the state is for sacrifice on the part of its members, and the solution of the evil lies in the analysis of the origins of the state. That such origins are founded in group conflicts, and that these group conflicts have *their* origin in the production process, is well recognized in modern social theory. The final solution must lie in the development of techniques for resolving class conflict, the principle of resolution to be based on a minimization of the basis for privileged groups through an increase in the productiveness of the society.

We assume, now, that meaning can ultimately be given to (i.e., progress can be made in the answering of) questions concerning individual interests, the degree of cooperation, and the degree of efficiency of a social group. These measures will provide the basis for specifications, production, and distribution of the manufactured article. There still remains the problem of the "control" of the manufacturing process. *If* we can specify the kind of material which will satisfy cooperative interests, *if* we can specify the amount and kind of material to be produced, *if* we can specify the manner of its distribution and use, we still require a method for deciding whether the specifications have been satisfied.

Now the matter of deciding whether a certain product meets the requirement of a specification would be quite simple to solve if all the material produced had exactly the same quality, or, more generally, if one could know without error the quality of any product. The decision as to whether or not a certain lot of material is acceptable would then depend solely upon the demands of the specification. For example, suppose we find that steel of a Rockwell hardness of exactly 67.2 is the best material out of which we can make a certain tool. If we could produce steel bars with exactly this Rockwell, the problem of deciding whether to use a lot of bars would be simple indeed.

Actually, of course, the situation is quite different. No manufacturing process can be refined to the point of producing exactly the same material over and over, just as no method of measurement can reproduce exactly the same observation (to any given decimal place).

Instead of demanding that all pieces have exactly the same quality, we might feel that the problem could be solved by demanding that the quality lie in a certain range. Thus, instead of an exact Rockwell of 67.2, we might require that the bars have hardnesses in the range $67.2 \pm .5$. But if this requirement is to be interpreted to mean that a controlled manufacturing process must *never* produce bars outside the range, then in general it will be impossible to meet the requirement for control. This is so because the statistical distribution of hardnesses, or other quality dimensions, even when we control the process to the extent of eliminating all systematic causes of fluctuation, can in general take on any reading if enough pieces are made. The probability that a given deviation from the mean will occur usually approaches zero the larger the deviation. But nevertheless in general a reading outside of any specified range *could* occur, even though control exists. This is a point so commonly overlooked in industrial practice that it has caused all kinds of confusion in the interpretation of specifications. In ordnance work, for example, it was supposed that if the specification required that no round of ammunition produce a chamber pressure exceeding 57,000 pounds per square inch, then this must mean that *absolutely* no round of a lot should produce a pressure in excess of this amount. Yet no one ever made a lot of ammunition for which there was no valid suspicion of a high pressure round. A specification so interpreted would lead us to reject all material, since the pressures of certain types of ammunition are apparently distributed normally around some mean value, and it is reasonable to expect that a very high value could occur, even though very rarely.

Thus we cannot say that a manufacturing process is "in control" only if all articles have exactly the same quality, or only if all articles have a quality measure within some finite range, since under such a definition we could not even begin to approach control. The situation is analogous to the control of observation, and remarks that we have already made on experimental control apply here as well. We must say that a certain manufacturing process is controlled provided a presupposition is made regarding the distribution of the quality of the pieces.

Suppose we consider first one possible method of determining

whether or not a process is controlled. The following information on the process and the sampling from lots is presupposed:

a) The separate lots produced are to be regarded as random samples from some normal universe with mean a_i and variance σ_i^2 with respect to the quality being measured.

b) The samples drawn from the lots to evaluate the quality are random samples.

c) The samples are all the same size (n) and are k in number.

Then to test whether the separate lots all have the same quality, i.e., whether all the means and variances are equal, one performs the following operations:

a) For each sample, compute the sample mean and variance, according to the formulae:

$$\bar{x}_i = \frac{\sum^n x}{n}$$

$$s_i^2 = \frac{\sum (x - \bar{x})^2}{n - 1}.$$

b) Over all the items compute

$$\bar{\bar{X}} = \text{grand average} = \frac{\sum^k \bar{x}_i}{k}$$

$$s_0^2 = \text{grand variance} = \frac{\sum (x - \bar{\bar{X}})^2}{nk - 1}.$$

c) Finally, compute

$$\log L_0 = \frac{1}{k} \sum^k \log s_i^2 - \log s_0^2.$$

d) Consult tables [1] of L_0 to determine the probability, P_0, that one will obtain values of L_0 in a certain range if the hypothesis of control is true. If P_0 is small, say .01, reject the hypothesis of control and take action to determine the reasons (e.g., maladjustment of machinery, tool-wear, etc.). If a value of P_0 of .01 or less is taken as a basis for action, then in the long run we will act mistakenly if a state of control of the presupposed type really exists, about one time in a hundred. Just how often on the basis of the above method

[1] See (3).

we will make the mistake of assuming control to exist, when in fact it does not, depends on the manner in which we are wrong. For example, the risks depend on the magnitude of the true differences in means and variances; if the true differences are small but not zero, the risks of mistakenly assuming control are large, and conversely, the risks of mistakenly assuming control become small when the true differences become large. The risks also depend on whether it is the means or the variances that differ, or whether the samples are random or from the same universe. From the practical viewpoint, it is the second kind of error (type II) that is most important; one must admit that rarely if ever will a true state of control exist, in the sense that all material produced will come from normal populations with the same means and variances. Statistical theory is only beginning to study the properties of these risks for a given procedure, but it cannot be denied that such research is highly important, in view of the practical results that depend on it.

Two important aspects of quality control procedures follow from this discussion, and indicate the fundamental problems involved in statistical control.

a) We have said that certain criteria are set up that enable us to decide methodologically whether or not a state of control exists. We have also pointed out that any given method will run the risk of being wrong and that this risk depends upon the "true" state of affairs, e.g., upon how much difference there is in mean quality from lot to lot. Nevertheless, we intend to take some action as a result of the control method we apply: either we intend to let the process run on as it has in the past, or else we intend to investigate it, and if necessary shut it down. Either action is risky if based on the wrong decision. If control does not exist, and we let the process go on, we are apt to pass defective material, while if acceptable control exists, and we stop the process to investigate, we are apt to waste labor on a useless activity. What criteria are we to use in prescribing the risks of control procedures? This question is extremely important, since the answer we give to it defines the kind of quality we can expect, and yet within present practice the answer is only given in an intuitive manner. Indeed, it is often difficult to convince industrial inspection departments that this question has real mean-

ing; it is supposed that a given inspection plan will satisfy *exactly* the demands of a specification, so that if material of a given strength, say, is required, we need only inspect to find out whether the items fulfill the requirement. Yet not even the much-vaunted one hundred per cent inspection will guarantee that the exacting demands of a specification are met, until we shall have removed all uncontrolled factors from inspection. Those of us who are familiar enough with the double risk of control procedures are nevertheless still puzzled as to how these risks can be determined. The last two chapters have attempted to formulate a basis for evaluating the consequences of the various types of wrong decision. We have pointed out that, for example, in the manufacture of certain drugs an excess of a certain compound may cause serious consequences to the users, whereas the cost of producing the item may not be large. Hence, we should endeavor to minimize to the greatest possible extent the risk of assuming control in the manufacturing process when control actually does not exist. We should even be willing to discard large quantities of the product if there is but slight evidence of poor material. The risk of discarding perfectly standard material might then be large, but the comparative cost of such a mistake is very small relative to the cost of issuing nonstandard material.

We have generalized upon this example, in order to determine the kind of information that is pertinent in setting risks. The ultimate purpose of production and distribution is to satisfy legitimate wants, we have said. The measure of value of a given production-distribution scheme is the degree to which it satisfies the maximum number of fundamental (historical) purposes. This degree is measured in terms of the probability that a given legitimate want will be satisfied over a group. There seems to be no reason why the risks of acceptance should not be based on this measure of efficiency of a production process; of all possible ways of setting the risks, that one is best which maximizes the measure of efficiency of a fixed production-distribution schedule.

This is facile defining, to be sure. It is simple enough, once one recognizes the possibility of measuring social efficiency, to present definitions of this kind. The real problem is one of deciding how to make the necessary measures. This is a problem for a future soci-

ology, but the main point of this discussion should be to direct effort towards such research, just as the main point of the entire essay is to set up criteria for adequate research. The thesis to be applied from the more general discussion is that sociological investigations will demand more than what observation plus statistics can give; they will demand a formalization of the structure of society, a "model," if you please, within which it is possible to formulate certain questions, and to measure the degree of progress in their answering.

Even though an experimental science of society has yet to be developed, the criteria we have given for setting risks does enable us to formulate a definition of "best control procedure" that can be applied roughly, since it is possible to make rough estimates of the social consequences of releasing products of a specified quality. The discussion has a moral for the statistician; it is too often supposed that the type of statistical test that is to be used can be decided on a priori grounds, by mathematical criteria only. The statistician sets down certain formal definitions of "best" test, which command more or less intuitive agreement. For example, we have already pointed out that there are certain statistical tests for control that have the following very strong property: for a given type I error (risk of rejecting a hypothesis when it is true), the test minimizes the type II error (error of accepting the hypothesis when it is false), regardless of the particular alternative hypothesis that happens to hold. Yet, even if it were obvious that a statistical test with this property is always preferable, nevertheless the conditions of its application are very restricted. Actually, we can hardly expect true control to exist, just as we can hardly expect two universe means to be exactly the same. Hence, the type I error practically never applies, and what we really want is a test that will reject control only when the lack of control is beyond some limit. But in such cases, the value criteria become very much more difficult to define. To the rationalist, or critical type mind, the answer to the question, which test is the best test to use, can be found within formal science, and this may be so if we give an arbitrary meaning to the concept of "best." But the writers on the subject evidently have in mind more than an arbitrary defining; they mean to imply that the use of the best statistical test will be the best course of action in accordance

with certain criteria of value. In this, they enter into the domain of value theory, and those who feel that purely formal answers can be given to what are the criteria of "best" are similar to the rationalist ethicists who feel that the ultimate criteria of good and bad action are to be found in intuitively given first principles.

No, the criteria of best test must be based on the general principles of a value theory. We have suggested here the fundamentals of a science of value. There can be little doubt that we must await considerable progress in social problems before this general value concept can be applied by a precise methodology to the problem of deciding which is the best statistical test. But there should be a realization in statisticians' minds that they have pushed their basic problem beyond the field of formal statistics when they attempt to set down the criteria of best test. The danger of not realizing this point lies in the possible action that will result when a formally defined criterion of best is taken to satisfy nonformal demands of the science of value. The function of the statistician is not to provide criteria for the best test, but rather to present a method for determining the chances of error associated with any given test, under any permissible hypothesis concerning the natural world.

b) Another aspect of control procedures, besides that of risks, is concerned with the presuppositions we make in establishing control. We usually assume that the sampling is random for a given lot. Is it? In the example given above, we assumed normality of the measures of quality. Was the assumption justified?

The natural reply to such questions would be to conduct certain tests to find out whether the assumptions hold. These tests themselves will be based on certain presuppositions, which for the moment we will have to grant, since no experiment can ever be run without some presupposition. But the true import of the problem still remains: we must start somehow in a given examination of control, and the question is to find the safest method of starting consistent with the demands of efficiency. In many cases, we can test for control by granting only randomness and a few very simple assumptions; in such cases, we are trying to determine whether the pieces are all being made under essentially the same conditions in the sense that, with respect to a given measure, they all come

from the same universe. Here we do not specify what the distribution function is that characterizes the universe, and hence escape the rather strong restrictions of normality. A Shewhart "control chart" (1), (2), and (4), is an example of this sort of test of control. As long as the items within any sample are taken at random, and the observations are independent, most of the averages should fall within certain tolerance limits, provided the original universe satisfies rather general conditions. Techniques other than the control chart are also available, and in some cases are preferable.

Even a general randomness need not be presupposed in testing for control, provided we can associate a specific distribution function with a given method of selecting items at a given time. For example, we may know that there is bound to be a downward trend in the mean measurement for a certain item over a period of time, and we may be able so to characterize this trend that at any time we can predict what the mean should be. We must presuppose, however, that randomness exists at any specified time.

Such is an outline of a method for determining whether control exists, but the method is not sufficient for practical purposes; the chief aim here is not merely to determine whether the manufacturing process is producing pieces from the same type of universe, but also to determine the limits within which a certain portion of the pieces produced should lie. The determination of these so-called tolerance limits is really the ultimate aim, but in general this aim cannot be attained without some assumption regarding the type of universe from which the pieces are drawn.

There has been some dispute in statistical literature on this problem of determining tolerance limits, the dispute mainly centering about the use of "small" or "large" samples. Objection has sometimes been made to the theory of tests of statistical hypotheses that was discussed in Chapter II; the objection is that the theory presupposes a "unique" sample, on the basis of which a decision is to be made. It is argued that in practice it would be foolhardy to trust the information gained from a small sample, even though the theoretical or mathematical background were exact. In particular, it is argued that the small sample can never be sufficient to establish tolerance limits.[1]

[1] See, for example, (5).

The actual dispute does not seem to have much force, though it undoubtedly is indirectly aimed at those who apply statistical theory with little or no regard for the presuppositions involved. Certainly, it is not meant that the size of the sample is to be thought of as a determining characteristic of experimental work, in the sense that all legitimate experiments must involve a minimum sample size. The argument of the present essay has been designed to show that both the observations and the formal presuppositions are the necessary postulates of any experiment; in some cases, we may so enrich the formal presuppositions that the number of observations required to make a decision is small, while in other cases we may use very simple presuppositions and require a large number of observations. Again, the formulation of the question will dictate the number of observations we will have to take. For example, if we so enrich the presuppositions as to include an assumption regarding the value of the variance (i.e., the "error"), as we may do in certain chemical analyses, then tolerance limits may be set by use of relatively few observations: all we would require would be a sufficient number of observations to establish our knowledge of the mean value within a certain degree of accuracy. We would then be able to predict that a certain fraction of the products will have values in a specified range. On the other hand, if we do not even presuppose that the universe from which the samples are drawn is normal, and presuppose nothing about the parameters of the universe, and prescribe randomness only for the statistics of samples, then in general a relatively large sample will be necessary in order to establish tolerance limits. The dispute concerning the correct procedure for setting up tolerance limits has certainly been an important one, for it has emphasized the role of presuppositions in statistical theory. It has shown those who are inclined to the mathematical side (the rationalists) that the formal presuppositions are critical, and that their justification is as important as the use of an exact statistical method; it has shown those who are inclined to the observational side (the empiricists) that there is no escape from some formal presuppositions, that data alone can never provide a basis for decision, and that the data themselves are presuppositions of a sort, the justification of which is as necessary as the justification of the formal presuppositions.

From this discussion we are led to a general definition of a state of statistical control relative to a manufacturing process: *A production process in which the number of pieces produced is unlimited is said to be a state of statistical control, provided we can assign (within measurable limits) a probability, associated with any random type of selection of the items, that any given measurement of a quality will lie within certain specified limits.*

Just what we shall presuppose in testing the truth of this statement with reference to any given process will depend upon the amount of information we can bring to bear on the problem, together with the risks associated with making an incorrect presupposition. If the evidence is plentiful that normality exists in the distribution of the observations, it will be foolish to employ weaker methods that do not presuppose normality, unless it is dangerous from the point of view of consumer interests to risk an incorrect distribution assumption. The generalized statistical purpose then becomes one of deciding upon the risks associated with any given type of procedure, provided the procedure happens to be based on a certain false assumption.

Before proceeding to a summary of the discussion of this chapter, a note of clarification should be added on the meaning of "control" as the term has been used here. In industry, the term "quality control" is usually applied to two aspects of the manufacturing procedure: process control and final control. In process control, we take action which stops the production or distribution process in order to investigate it, if we have reason to suspect that control no longer exists. In final control, we "discard" lots of material on the same basis. In this discussion, we have been using the term control to include either sense; which kind of control is preferable, process or final (or both), will depend upon considerations of efficiency.

The theme of the chapter has been one of defining the meaning of productive human activity, by assigning to it a generalized purpose. We left unanswered, for the moment, the ultimate purpose of science, in order to describe the purpose and procedures of the scientific analogue: production and distribution. We saw that progress in the domain of production and distribution demanded (a) a measure of consumer interest, and, in particular, a measure of

cooperative interests; (b) a measure of social efficiency, in terms of the maximum of satisfaction of legitimate desires; (c) a measure of the degree to which a manufacturing process can be said to satisfy the specifications that follow from a) and b).

That science is a necessary condition for the approach to the ideals entailed in these three activities is apparent enough, and apparent also is the value of the scientific ideal. The whole may be concisely summarized and concluded in Singer's phrase: "The measure of man's cooperation with man in the conquest of nature measures progress" (6).

REFERENCES

1. *A.S.T.M. Manual on Presentation of Data*, American Society for Testing Materials, Philadelphia (1943).
2. *Control Chart Method of Controlling Quality of Manufactured Product*, A.S.A. War Standard Z 1.3.
3. Freeman, H., *Industrial Statistics*, John Wiley (1942).
4. Shewhart, W. A., *Economic Control of Quality of Manufactured Product*, Van Nostrand (1941).
5. Shewhart, W. A., *Statistical Method from the Viewpoint of Quality Control*, Graduate School, U. S. Dept. of Agriculture (1939).
6. Singer, E. A., "On Progress," in *The Contented Life*, Holt (1936).

Index